PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/12, 15/11, 1/15, 1/21, 5/10, 15/62, C12Q 1/68, C07K 14/47, 16/18, A61K 31/70, 38/17, G01N 33/68, 33/574 // (C12N 1/15, C12R 1:84) (C12N 1/21, C12R 1:19)

A2

(11) International Publication Number: **WO 99/47671**

(43) International Publication Date: 23 September 1999 (23.09.99)

(21) International Application Number: PCT/IB99/00495

(22) International Filing Date: 12 March 1999 (12.03.99)

(30) Priority Data:
60/077,997 13 March 1998 (13.03.98) US

(71)(72) Applicants and Inventors: LUKANIDIN, Eugene M. [RU/DK]; Willemoesgade 33, 1th, DK–2100 Copenhagen O. (DK). CHRISTENSEN, Clause, R., L. [DK/DK]; Gammel Kongevej 35 – 4TV, DK–1610 Copenhagen S.W. (DK).

(74) Agent: KOLB, Helga; Hoffmann . Eitle, Arabellastrasse 4, D–81925 München (DE).

(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: DNA SEQUENCES ENCODING SEMAPHORIN–H AND DIAGNOSIS OF METASTATIC CANCER

(57) Abstract

The present invention relates to DNA molecules encoding novel members of the Semaphorin family, referred to as Semaphorin–H and Semaphorin–H–v. The invention further relates to polypeptides encoded by the DNA, to antibodies that bind to the polypeptides and to compositions and methods comprising the DNA, polypeptides and antibodies. The present invention provides methods for detecting metastatic cancer and determining the metastatic potential of cells. The invention contemplates the development of cells line with varying metastatic potential for use a models for studying, diagnosing and treating metastatic disease.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

DNA SEQUENCES ENCODING SEMA-H AND DIAGNOSIS OF METASTATIC CANCER

TECHNICAL FIELD OF THE INVENTION

The invention relates to DNA molecules comprising sequences encoding a novel member of the Semaphorin family, Semaphorin-H. The invention further relates to the polypeptides encoded by those DNA sequences and antibodies that bind to those polypeptides. The invention also relates to recombinant DNA molecules comprising these DNA sequences, as well as hosts transformed with such recombinant DNA molecules. The invention further relates to the use of the DNA molecules, polypeptides and antibodies of this invention in the study, diagnosis and treatment of metastasis.

The present invention further provides methods for the diagnosis of metastatic cancer by detection of the sema-H mRNA or the sema-H protein encoded by the sema-H gene. The present invention contemplates the use of recombinant sema-H DNA and antibodies directed against the sema-H protein to determine the metastatic potential of several types of tumor cells, including, for example, thyroid, epithelial, lung, liver, kidney, breast, lymphoid, hematopoietic, pancreatic, endometrial, ovarian, cervical, skin, colon and similar tumor cells.

The present invention also provides mammalian cell lines and tumors with high and low metastatic potential developed through use of the disclosed invention. Such cell lines to serve as useful model systems for in vitro and in vivo anti-metastasis drug screening.

systems for in vitro and in vivo anti-metastasis drug screening.

BACKGROUND OF THE INVENTION

Malignant cancer tumors shed cells which migrate to new tissues and create secondary tumors; a benign tumor does not generate secondary tumors. The process of generating secondary tumors is called metastasis and is a complex process in which tumor cells colonize sites distant from the primary tumor. Tumor metastasis remains the major cause of morbidity and death for patients with cancer. One of the greatest challenges in cancer research is to understand the basis of metastasis, i.e., what controls the spread of tumor cells through the blood and lymphatic systems and what allows tumor cells to populate and flourish in new locations.

The metastatic process appears to be sequential and selective, and is controlled by a series of steps since metastatic tumor cells: (a) are mobile and can disseminate from the original tumor; (b) are capable of invading the cellular matrix and penetrating through blood vessels; (c) possess immunological markers, which allow them to survive passage through the blood stream, where they must avoid the immunologically active cytotoxic "T" lymphocytes; and (d) have the ability to find a favorable location to transplant themselves and successfully survive and grow.

Understanding the underlying molecular mechanisms in metastasis is the key to understanding cancer biology and its therapy. In clinical lesions,

malignant tumors contain a heterogeneous population of cells, exhibiting a variety of biological characteristics, e.g., differential growth rates, cell surface structures, invasive capacities and sensitivity to various cytotoxic drugs. Researchers can take advantage of tumor heterogeneity factors, by identifying specific cell produced markers, which are unique for metastasis, to develop therapeutic regimens which do not rely only on surgical resection alone.

At this time it is not known whether the metastatic phenotype is under the regulation of a single gene or multiple genes ,which genes may be independent or interrelated. However, a number of genes have become correlated with the formation and metastasis of tumors. For example, several normal cellular genes become oncogenes by incorporation into a retroviral genome. Due to the juxtaposition of new promoter elements, such incorporation frequently allows potential oncogene to be expressed in inappropriate tissues or at higher levels than it normally would be expressed. It appears from work with tumorigenic retroviruses as well as other systems, that misexpression of many cellular proteins, particularly those involved in the regulation of the cell cycle, cell mobility, or cell-cell interaction may lead to a cancerous phenotype. It is, therefore, important to identify genes involved in the biological pathways of metastasis as this identification is critical to the development of preventative, diagnostic and treatment measures.

The Semaphorin/Collapsin family of molecules was discovered recently and is characterized by unique and highly conserved motifs within a 500 amino acid semaphorin domain. Proteins of the family all contain

a signal sequence and are either transmembrane or secreted. Grasshopper SemaI, Drosophila SemaII, mouse SemaD, chicken collapsinI and mouse SemaIII exhibit inhibitory or repulsive functions in a neuronal context. However, the wide expression of some semaphorin members indicates alternative functions for the proteins. SemaIII knockout mice, for example, show abnormal bone structure, in addition to neural abnormalities.

Data also implicate semaphorins in the pathogenesis of disease. A semaphorin domain is encoded by variola, vaccinia and acelaphine herpes virus 1. The sema IV and semaV genes are located in the 3p21.3 chromosomal region in humans, a region that is deleted in many small cell lung cancer tumors.

The present invention discloses the mouse sema-H gene and its usefulness in the diagnosis of metastatic cancer by use of either antibodies directed against the sema-H protein or sema-H nucleic acid probes directed against sema-H mRNA.

SUMMARY OF THE INVENTION

The present invention provides semaH polypeptides, and fragments and derivatives thereof as well as nucleic acids encoding them. Another aspect of the invention provides antibodies that specifically bind to semaH polypeptides of the invention. A further aspect of the invention provides antisense oligonucleotides derived from the nucleic acid sequences of the invention. The present invention also provides methods and compositions for detecting and preventing metastasis utilizing the polypeptides,

nucleic acids, antisense oligonucleotides and antibodies of the invention.

Additionally, the present invention provides recombinant DNA molecules comprising the DNA sequences of the invention, host cells transformed with the recombinant DNA molecules and methods for producing the polypeptides encoded by those DNA sequences utilizing the transformed host cells.

Therefore, one aspect of the present invention provides an isolated, recombinant nucleic acid encoding a human Semaphorin-H gene or a fragment thereof, and replicable DNA sequences encoding a Semaphorin-H polypeptide which express high or low levels of the Semaphorin-H polypeptide. Isolated antisense Semaphorin-H nucleic acids and expression vectors are also contemplated by the present invention. Human Semaphorin-H nucleic acids are preferred. The DNA sequences of this invention may be used in both therapeutic and diagnostic applications. The DNA sequences encoding inactive, mutant or truncated forms of Semaphorin-H are useful in gene therapy to prevent metastasis of cancer cells by competing with the native, active form of the protein in the cell.

A further aspect of this invention provides isolated transformed host cells, such as prokaryotic microorganisms, yeast, insect cells and eukaryotic cells, containing Semaphorin-H nucleic acids and replicable vectors containing DNA sequences encoding the Semaphorin-H polypeptide.

A still further aspect of this invention provides isolated mammalian Semaphorin-H polypeptides and pharmaceutical compositions comprising them. Human Semaphorin-H polypeptides are preferred.

The DNA sequences encoding Semaphorin-H are suitable for use in creating transformed cell lines with high metastatic potential, especially in cancer cell lines, for example, for use in research.

The DNA sequences of this invention may be used as diagnostic tools to detect and quantify Semaphorin-H mRNA levels in various cells. This method may be used to determine the metastatic potential of cancer cells, a high Semaphorin-H mRNA level suggesting that the cell has high metastatic potential.

The Semaphorin-H DNA sequences of this invention may also be used to inhibit the expression of Semaphorin-H polypeptides in a cell through the use of anti-sense technology. Single stranded, anti-sense DNA can be introduced into cells where it can hybridize to and inhibit the translation of Semaphorin-H mRNA. Such methods can be used to prevent a cancer cell from metastasizing.

The invention also provides monoclonal and polyclonal antibodies directed against a Semaphorin-H polypeptide or any peptide, fragment or derivative of the Semaphorin-H protein. These antibodies may be used to assay for Semaphorin-H levels in a patient or cell. In addition, Semaphorin-H antibodies are useful in inhibiting Semaphorin-H activity, and therefore, metastasis. One aspect of the present invention is directed to a method for diagnosing metastatic cancer by contacting serum from an individual to be tested for such cancer with an antibody reactive with a mammalian Semaphorin-H protein or an antigenic fragment thereof, for a time and under conditions sufficient to form an antigen-antibody complex, and detecting the antigen-antibody complex.

A further aspect of this invention is directed towards treatment of cancer by administering reagents, such as for example, anti-Semaphorin-H antibodies capable of binding the Semaphorin-H protein and antisense Semaphorin-H nucleic acids capable of binding Semaphorin-H sense mRNA.

Yet another aspect of the present invention provides an animal model system of the metastatic process, including several eukaryotic cell lines and tumors expressing different levels of Semaphorin-H, which can be derived, for example, from mouse and rat carcinomas. These cell lines and tumors may be re-introduced, for example, into mice or rats to produce primary tumors which metastasize to the lung, liver and kidneys with a characteristic frequency. Therefore, the present invention also provides a well controlled animal model system for testing pharmaceutical compositions suspected to have therapeutic utility for the treatment of metastatic cancer.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1: Identification of M-semaH transcripts in metastatic cell lines

(A) Differential display showing the amplification of a 525 bp fragment from the two metastatic cell lines 66c14 and 4T1 (arrow).

(B) Confirmation of the differential expression by Northern hybridization of the 525 bp fragment to RNA from tumor cell lines. 10 µg total RNA was loaded in each lane.

(C) Northern hybridization of the 525 bp fragment to various mouse tumor cell lines. Note the degree of correlation with metastatic potential in figures B and

C as well as the presence of 4.0 kb, 4.5 kb and 7.0 kb transcripts.

[a]VMR-Ly was found to be metastatic to the lymph nodes but not the lungs. The presence of transcripts was confirmed by prolonged exposition (data not shown).
[b]Not determined.
S.C.= spontaneous metastasis assay, injection in the subcutis; i.v. = experimental metastasis assay, injection intravenously; +=metastatic; -= nonmetastatic.

Figure 2: The difference between the 4.5 kb and 4.0 transcripts

(A) Diagram depicting the 4466 bp and the 3989 bp cDNA fragments obtained by library screening. Binding sites for different probes (A, B, C, D) are indicated above. Probe A corresponds to the 525 bp fragments obtained from the display in Figure 1A. Arrows indicate the position of primers in the 5' RACE. The interleaved lines indicate the additional sequences retrieved from the 5' RACE.

(B) Northern hybridization of probe B to RNA from tumor cell lines 4T1 (lane 1), 66c14 (lane 2), 4T07 (lane 3).

(C) Northern hybridization using probe C to the same RNA as in (B). Note that the region missing in the 3675 bp cDNA fragment is also absent in the 4.0 kb *M-sema H-v* transcript.

(D) Northern hybridization with probe D to the same RNA as in (B). The presence of the 4.0 kb *M-sema H-v* transcript shows that this transcript shares the 5'-end with the 4.5 kb transcript.

Figure 3: Expression of *M-semaH* transcripts in mouse adult tissues and whole embryos.

(A) Upper: Northern hybridization to total RNA from tissues of A/Sn mice using probe A (see Figure 2A). Note the absence of expression in normal mammary tissue. Lower: control using a GAPDH probe.

(B) Upper: Norther hybridization to total RNA of whole A/Sn embryos between 10.5-18.5 dpc of mouse development as well as from the head and back of a 15.5 dpc embryo and the brain of a newborn mouse. Same probe as in (A). Note the expression before day 13.5. Lower: control using ^{32}P-labeled poly(U). (A and B) Both filters were exposed for one week. The position of the 4.0 kb transcript in the 66c14 lanes is recognizable from a 24 h exposure placed next to the filter and marked with an asterisk.

Abbreviations: B, back; H, head; NB, newborn; dpc, day(s) post coitum.

Figure 4: Conserved motifs in the predicted M-SemaH sequence

(A) Alignment of the Sema domains of M-SemaH, C-Coll 5, H-Sema III, D-SemaII, and G-SemaI. The alignment was made using the Pileup program (Genetics Computer Group). Conserved cysteine residues are marked with asterisks. Potential N-glycosylation sites in M-SemaH are marked with number symbols. Conserved amino acids are boxed.

(B) Diagram depicting the organization of the proteins aligned in (A). The percentages refer to the degree of identity between the entire amino-acid sequence of M-SemaH and the proteins shown as calculated using the

Gap program (Genetics Computer Group). A small black bar in the N-terminal means that a signal peptide is present. A plus in the C-terminal means that 18-20 positively charges residues are present within the last 50 aa. Sema, Semaphorin; Ig, immunoglobulin-like domain; TM, transmembrane domain. M, mouse; C, chick; Coll, Collapsin; H, human; D, *drosophila*; G, Grasshopper.

Figure 5: *In situ* hybridization analysis of *M-semaH* expression during mouse development.

Darkfield illuminations of hybridizations with an *M-sema H/H-v* antisense probe (A, C, E, G) Brightfield illuminations of sections succeeding A, C, E, G stained with haematoxylin-eosin (he) (B, D, F, H). (A-D) Transversal section through the body of a 12.5 dpc embryo.

(A) The *M-semaH* transcripts are expressed in the ventral horns (arrow heads) of the neural tube and in sclerotomal cells underlying the spinal ganglia.

(B) A section following (A) stained with He.

(C) The section succeeding (A) showing sclerotomal signals but no signals in the ventral horns.

(D) He-staining of section succeeding (C).

(E) Parasagittal section of the lower body of a 14 dpc embryo. The *M-semaH* signals are visible in the developing intervertebral discs.

(F) He-staining of the section succeeding E. Arrow heads in (E) and (F) indicate the positions of the intervertebral discs.

(G) Sagittal section through the lung of a 16 dpc embryo. The *M-semaH* transcripts are detectable in the epithelium of the bronchi.

(H) He-staining of a section succeeding G.

Bv, blood vessel; br, bronchus, id, primordium of the intervertebral discs; nt, neural tube; sg, spinal ganglia; vb, primordium of the vertebral body.

DETAILED DESCRIPTION OF THE INVENTION

The following definitions are used throughout the specification and the claims.

"Sema-H" refers to nucleotide sequences, either DNA or RNA, single-stranded or double-stranded that encode a Sema-H polypeptide or derivative thereof. When used to define a double-stranded nucleotide sequence, the term also refers to the anti-sense nucleotide sequence.

"Sema-H polypeptide" refers to a polypeptide which is at least 80%, and more preferably, at least 90% identical in amino acids 29-303 of SEQ ID NO:3 or SEQ ID NO:4.

Additional terms are defined where necessary throughout the application.

According to one embodiment, the invention provides isolated DNA molecules comprising the DNA sequences of SEQ ID NO:1 and SEQ ID NO:2 or fragments thereof; DNA sequences that hybridize under stringent conditions to SEQ ID NO:1 and SEQ ID NO:2 or fragments thereof; DNA sequences which encode a polypeptide having the same amino acid sequences encoded by SEQ ID NO:3 or SEQ ID NO:4 or a fragment thereof; and sequences which hybridize thereto under stringent conditions.

As used herein, the term "hybridize to under stringent conditions" refers to the ability of a

denatured DNA sequence to hydrogen bond to another denatured DNA sequence through complementary base pairs under conditions which allow sequences having at least 80% similarity to form such hybrids. Such conditions are well known in the art and are exemplified by salt and temperature conditions substantially equivalent to 5X SSC and 65° for both hybridization and wash.

Besides the full-length cDNA sequences set forth herein, it will be readily apparent to those of skill in the art that any other DNA sequence which, as a result of degeneracy in the genetic code, encodes the same amino acid sequence as SEQ ID NO:3 or SEQ ID NO:4 is part of applicant's invention. While those specific sequences are not set forth herein due to space considerations, it should be understood that one of ordinary skill in the art could ascertain all of such DNA sequences merely by reference to the genetic code and without the exercise of inventive skill.

In addition to genetically redundant DNA sequences, the invention also includes DNA sequences which encode other amino acid sequences which are at least 80%, and preferably at least 90% similar to SEQ ID NO:3 or SEQ ID NO:4. Because this aspect of the invention does not require the isolated DNA sequence to encode an active Sema-H, any nucleotides of SEQ ID NO:1 and SEQ ID NO:2 may be modified to produce a DNA sequence of this invention. The identification and isolation of additional sequences may be achieved by standard DNA library screening techniques (hybridization, PCR) using SEQ ID NO:1 and SEQ ID NO:2 or portions thereof as a probe. Such homologous sequences may be found in any mammalian tissue cDNA library, as well as in insect cDNA libraries, yeast and

other fungi cDNA libraries and prokaryotic cDNA libraries.

According to an alternate embodiment, the invention provides for isolated DNA sequences which encode Sema-H polypeptides; DNA sequences which hybridize to either of the former DNA sequences; and DNA sequences which code for a polypeptide having the same amino acid sequence as any of the previous DNA sequences.

Additional DNA sequences which encode Sema-H polypeptides may be identified by standard DNA library screening techniques using nucleotides of SEQ ID NO:1 and SEQ ID NO:2 or portions thereof as a probe. Even more preferred are homologous DNA sequences which contain nucleotides encoding only conservative amino acid substitutions for some or all of the other amino acids of SEQ ID NO:3 or SEQ ID NO:4. The translation products of any of these DNA sequences may then be used in assays or methods as described below.

As set forth above, DNA sequences according to this aspect of the invention may be identified and isolated using methods well known in the art, for example, through standard cDNA library screening. It will be appreciated by one of skill in the art that screening techniques such as those described in the art may be used to identify homologous genes in other species, and these DNA and amino acid sequences are also included in the present invention.

According to another embodiment of this invention, peptide fragments of Sema-H can be generated from the full length sequence using chemical or recombinant DNA techniques. Alternatively, or in combination, synthetic fragments of DNA or protein which represent portions of the Sema-H sequences

identified above may be generated. In a preferred embodiment, these fragments represent unique portions of Sema-H in that they are not found in any previously known protein. A unique portion would preferably have an amino acid sequence length at least long enough to define a novel peptide. Depending on the particular amino acid sequence, this unique portion would preferably consist of about 5 to about 25 amino acids, or most preferably about 5 to about 10 amino acids. These unique portions can be identified by comparing the amino acid sequence of Sema-H with known data base sequences.

As would be known to those skilled in the art, these unique portions, or any of the sequences named herein, may be free or coupled to other atoms or molecules, or they may be contiguous with a larger polypeptide derived from any source. They may be modified or joined to other compounds using any technique known to those skilled in the art, including but not limited to, physical, chemical or molecular techniques. These modifications may affect properties of the Sema-H or Sema-H derived polypeptides in a manner predictable to those with skill in the art. These properties may include solubility, stability, binding specificity, affinity, toxicity, localization, detectability, half life, targeting, bioavailability, antibody reactivity, protein folding, etc. and would be readily identifiable using appropriate assays.

According to another embodiment of this invention, any of the DNA sequences described above may be employed in a recombinant DNA molecule. The isolated DNA sequences of this invention may be inserted into any of the numerous commercially or otherwise publicly available cloning vectors. Useful

expression vectors, for example, may consist of segments of chromosomal, nonchromosomal or synthetic DNA sequences. Suitable vectors include, but are not limited to, derivatives of SV40 and known bacterial plasmids, e.g., E.coli plasmids col El, pCR1, pBR322, pMB9 and their derivatives, plasmids such as RP4; phage DNAs, e.g., the numerous derivatives of phage λ, e.g., NM989, and other phage DNA, e.g., M13 and Filamentous single stranded phage DNA; yeast plasmids such as the 2μ plasmid or derivatives thereof; vectors useful in eukaryotic cells, such as vectors useful in insect or mammalian cells; vectors derived from combinations of plasmids and phage DNAs, such as plasmids that have been modified to employ phage DNA or other expression control sequences; and the like.

The choice of cloning vector will, of course, depend upon the cell type(s) to be transformed with the resulting recombinant DNA molecule, as well as the ultimate use of that recombinant DNA molecules (i.e., DNA production, gene therapy, polypeptide expression). Preferred eukaryotic vectors are SV40-derived vectors. Preferred prokaryotic vectors are *E. coli* expression vectors. Preferred viral vectors are modified eukaryotic viral vectors, preferably attenuated adenovirus vectors, that may be used in gene therapy, such as those described in PCT publications WO 94/26915 and WO 94/28938, the disclosures of which are herein incorporated by reference.

Techniques for inserting a DNA sequence of this invention into a vector to produce a recombinant DNA molecule involve standard molecular biological techniques and are well known in the art (see, for example, J. Sambrook, et al., Molecular Cloning, A

Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)).

According to a preferred embodiment, the recombinant DNA molecule of this invention will additionally comprise an expression control sequence operatively linked to a DNA sequence of this invention. The term "expression control sequence" refers to a DNA sequence that controls and regulates the transcription and translation of another DNA sequence. It includes both 5' and 3' non-coding DNA sequences and optionally includes an ATG start codon.

The term "operatively linked" refers to the positioning of an expression control sequence with respect to a coding DNA sequence of interest such that the expression control sequence controls and regulates the transcription and translation of that DNA sequence. The term "operatively linked" includes having an appropriate start signal (e.g., ATG) in front of the DNA sequence to be expressed and maintaining the correct reading frame to permit expression of the DNA sequence under the control of the expression control sequence and production of the desired product encoded by the DNA sequence. If a gene or DNA sequence that one desires to insert into a recombinant DNA molecule does not contain an appropriate start signal, such a start signal can be inserted in front of the gene.

The choice of expression control sequence depends upon the nature of the recombinant DNA molecule, the host that will be transformed by that recombinant DNA molecule, and whether constitutive or inducible expression of a DNA sequence of this invention is desired. Such useful expression control sequences include, for example, the early and late promoters of SV40 or adenovirus, the lac system, the

trp system, the TAC or TRC system, the major operator and promoter regions of phage λ, the control regions of fd coat protein, the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase (e.g., Pho5), the promoters of the yeast α-mating factors, and other sequences known to control the expression of genes of prokaryotic or eukaryotic cells or their viruses, and various combinations thereof.

Preferably the vector also encodes a selectable marker, for example, antibiotic resistance. Replicable expression vectors can be plasmids, bacteriophages, cosmids and viruses. Any expression vector comprising RNA is also contemplated.

Preferred vectors of the present invention are derived from eukaryotic sources. Expression vectors that function in tissue culture cells are especially useful, but yeast vectors are also contemplated. These vectors include yeast plasmids and minichromosomes, retrovirus vectors, BPV (bovine papilloma virus) vectors, baculovirus vectors, SV40 based vectors and other viral vectors. SV40-based vectors and retrovirus vectors (e.g., murine leukemia viral vectors) are preferred. Tissue culture cells that are used with eukaryotic replicable expression vectors include Sf21 cells, CV-1 cells, COS-1 cells, NIH3T3 cells, mouse L cells, HeLa cells and such other cultured cell lines known to one skilled in the art.

A baculovirus expression system can be used to produce large amounts of Sema-H polypeptides in cultured insect cells. The post-translational processing-of polypeptides produced in such insect cells is similar to that of mammalian calls. Production of polypeptides in insects is therefore

advantageous, particularly when one seeks to mimic the exact function or antigenic properties of the natural polypeptide.

Methods for producing polypeptides in the baculovirus expression system are known to the skilled artisan. See for example Miller 1988 Ann. Rev. Microbiol. 42: 177. In general, a modified Autographa californica nuclear polyhedrosis virus propagated in Sf21 cells is used for polypeptide expression. This modified virus is produced by cotransfection of a small transfer vector, encoding a Sema-H polypeptide, with a viral expression vector which has been linearized within an essential gene. Once inside the cell, the linearized expression vector can undergo recombination with the transfer vector or simply recircularize. However, only recombination gives rise to viable viruses because the function of the essential gene is lost by recircularization. Recombinant expression viruses are detected by formation of plaques.

The present invention also contemplates prokaryotic vectors that may be suitable for expression of the mammalian Sema-H gene, including bacterial and bacteriophage vectors that can transform such hosts as E. coli, B. subtilis, Streptomyces sps. and other microorganisms. Many of these vectors are based on pBR322 including Bluescript™ (commercially available from Stratagene) and are well known in the art. Bacteriophage vectors that are used in the invention include lambda and M13.

Sequence elements capable of effecting expression of the Sema-H gene include promoters, enhancer elements, transcription termination signals and polyadenylation sites. Promoters are DNA sequence elements for controlling gene expression, in

particular, they specify transcription initiation sites. Prokaryotic promoters that are useful include the lac promoter, the trp promoter, and P_L and P_n promoters of lambda and the T7 polymerase promoter. Eukaryotic promoters are especially useful in the invention and include promoters of viral origin, such as the SV40 late promoter and the Moloney Leukemia virus LTR, Murine Sarcoma Virus (MSV) LTR, yeast promoters and any promoters or variations of promoters designed to control gene expression, including genetically-engineered promoters. Control of gene expression includes the ability to regulate a gene both positively and negatively (i.e., turning gene expression on or off) to obtain the desired level of expression.

The replicable expression vectors of the present invention can be made by ligating part or all of the Sema-H coding region in the sense or antisense orientation to the promoter and other sequence elements being used to control gene expression. This juxta-positioning of promoter and other sequence elements with the Sema-H gene allows the production of large amounts of sense or antisense Sema-H mRNA. Large amounts of the Sema-H protein can also be produced which are useful not only for anti-Sema-H antibody production but also for analysis of the function of Sema-H in metastatic cancer as well as for designing therapies for metastatic cancer. Analysis of the Sema-H protein, in this embodiment, includes the use of recombinant protein for use in crystallographic examination. Crystallographic examination of Sema-H could be used to identify small molecule inhibitors of active sites on the protein.

Therefore, one skilled in the art has available many choices of replicable expression vectors, compatible hosts and well-known methods for making and using the vectors. Recombinant DNA methods are found in any of the standard laboratory manuals on genetic engineering.

The invention also provides host cells transformed by the recombinant DNA molecules of this invention. These hosts may include well known eukaryotic and prokaryotic hosts, such as strains of E.coli, Pseudomonas, Bacillus, Streptomyces, fungi such as yeasts, and animal cells, such as WHO, RLL, B-W and L-M cells, African Green Monkey kidney cells (e.g., COS 1, COS 7, BSC1, BSC40, and BMT10), insect cells (e.g., Sf9), and human cells and plant cells in tissue culture. Eukaryotic cells may harbor the recombinant DNA molecules of this invention as an extra chromosomal element or incorporate all or part of it into the host chromosome.

It will be understood that not all vectors, expression control sequences and hosts will function equally well to replicate and/or express the DNA sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one skilled in the art will be able to select the proper vectors, expression control sequences, and hosts without undue experimentation to accomplish the desired expression without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must function in it. The vector's copy number, the ability to control that copy number, and the expression of any other proteins encoded by the vector, such as

antibiotic and other selective markers, will also be considered.

In selecting an expression control sequence, a variety of factors will normally be considered. These include, for example, the relative strength of the system, its controllability, and its compatibility with the particular DNA sequence or gene to be expressed, particularly as regards potential secondary structures. Suitable unicellular hosts will be selected by consideration of, e.g., their compatibility with the chosen vector, their ability to fold proteins correctly, and their fermentation requirements, as well as the toxicity to the host of the product encoded by the DNA sequences to be expressed, and the ease of purification of the expression products.

Considering these and other factors a person skilled in the art will be able to construct a variety of vector/expression control sequence/host combinations that will replicate and/or express the DNA sequences of this invention on fermentation or in large scale animal culture.

Methods for transforming cells with recombinant DNA molecules are well known in the art (see, for example, J. Sambrook, et al., Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1989)). Any of those methods may be employed to produce the transformed hosts of this invention. Identification of transformed hosts may be achieved by assaying for the presence of Sema-H DNA, Sema-H RNA or Sema-H polypeptide. Additionally, transformants may be identified by growth in selective media. For this assay, the gene necessary for growth in selective media is cotransfected into the cell either on the same or a

different recombinant DNA molecule as the Sema-H DNA and is expressible in that cell. It will, of course, be obvious that the gene for selective growth should not be present in the untransformed cell.

The transformed hosts of this invention may be employed to produce either large quantities of Sema-H DNA sequences of this invention and/or the Sema-H polypeptide encoded thereby. In order to produce large quantities of Sema-H DNA, the host, preferably a prokaryotic host, is grown in a medium and under conditions that promote DNA replication and cell division. Any complete media routinely used to grow bacteria is suitable for this purpose. Following growth, the transformed cells are separated from the growth medium and plasmid DNA is then isolated from the cells by standard and well-known techniques. Sema-H DNA may then be excised from the plasmid through the use of restriction endonucleases.

When the transformed hosts of this embodiment are employed, the preferred host is a mammalian cell. The transformed host should be grown in a medium that promotes expression of the Sema-H polypeptide-encoding DNA sequence present in that host. If expression of that DNA sequence is under the control of a constitutive promoter, any standard growth medium is suitable. If the Sema-H DNA is under the control of an inducible promoter, the growth medium should be supplemented with a compound that induces expression or growth conditions should be altered so as to induce expression (i.e., change in growth temperature). Following expression, the transformed cells are separated from the growth medium, lysed and the Sema-H polypeptide is purified by standard methods. If the cells secrete the Sema-H polypeptide, the protein may

be harvested directly from the media without cell lysis.

In a further embodiment, the present invention further provides derivatives of the Sema-H polypeptide. As used herein, a "derivative" of a Sema-H polypeptide according to his invention is a Sema-H polypeptide in which one or more physical, chemical, or biological properties has been altered. Such modifications include, but are not limited to: amino acid substitutions, modifications, additions or deletions; alterations in the pattern of lipidation, glycosylation or phosphorylation; reactions of free amino, carboxyl, or hydroxyl side groups of the amino acid residues present in the polypeptide with other organic and non-organic molecules; and other modifications, any of which may result in changes in primary, secondary or tertiary structure.

In accordance with this invention, derivatives of the novel Sema-H polypeptides may be prepared by a variety of methods, including by *in vitro* manipulation of the DNA encoding the native polypeptides and subsequent expression of the modified DNA, by chemical synthesis of derivatized DNA sequences, or by chemical or biological manipulation of expressed amino acid sequences.

For example, derivatives may be produced by substitution of one or more amino acids with a different natural amino acid, an amino acid derivative or non-native amino acid, conservative substitution being preferred, e.g., 3-methyl histidine may be substituted for histidine, 4-hydroxyproline may be substituted for proline, 5-hydroxylysine may be substituted for lysine, and the like.

Causing amino acid substitutions which are less conservative may also result in desired derivatives, e.g., by causing changes in charge, conformation and other biological properties. Such substitutions would include for example, substitution of a hydrophilic residue for a hydrophobic residue, substitution of a cysteine or proline for another residue, substitution of a residue having a small side chain for a residue having a bulky side chain or substitution of a residue having a net positive charge for a residue having a net negative charge. When the result of a given substitution cannot be predicted with certainty, the derivatives may be readily assayed according to the methods disclosed herein to determine the presence or absence of the desired characteristics.

In a preferred embodiment of this invention, the Sema-H polypeptides disclosed herein are prepared as part of a larger fusion protein. For example, a Sema-H polypeptide of this invention may be fused at its N-terminus or C-terminus to a different immunogenic Sema-H polypeptide, to a non-Sema-H polypeptide or to combinations thereof, to produce fusion proteins comprising the Sema-H polypeptide.

The Sema-H polypeptides may also be part of larger multimeric molecules which may be produced recombinantly or may be synthesized chemically. Such multimers may also include the polypeptides fused or coupled to moieties other than amino acids, including lipids and carbohydrates.

It will be readily appreciated by one of ordinary skill in the art that the Sema-H polypeptides of this invention, as well as fusion proteins and multimeric proteins containing them, may be prepared by

recombinant means, chemical means, or combinations thereof.

The molecules comprising the Sema-H polypeptides encoded by the DNA sequences of this invention, or derivatives thereof, may be isolated from the fermentation or cell culture and purified using any of a variety of conventional methods including: liquid chromatography such as normal or reversed phase, using HPLC, FPLC and the like; affinity chromatography (such as with inorganic ligands or monoclonal antibodies); size exclusion chromatography; immobilized metal chelate chromatography; gel electrophoresis; and the like. One of skill in the art may select the most appropriate isolation and purification techniques without departing from the scope of this invention.

In addition, the Sema-H polypeptides may be generated by any of several chemical techniques. For example, they may be prepared using the solid-phase synthetic technique originally described by R. B. Merrifield, "Solid Phase Peptide Synthesis. I. The Synthesis Of A Tetrapeptide", J. Am. Chem. Soc., 83, pp. 2149-54 (1963), or they may be prepared by synthesis in solution. A summary of peptide synthesis techniques may be found in E. Gross & H. J. Meinhofer, 4 The Peptides: Analysis, Synthesis, Biology; Modern Techniques Of Peptide And Amino Acid Analysis, John Wiley & Sons, (1981) and M. Bodanszky, Principles Of Peptide Synthesis, Springer-Verlag (1984).

Typically, these synthetic methods comprise the sequential addition of one or more amino acid residues to a growing peptide chain. Often peptide coupling agents are used to facilitate this reaction. For a recitation of peptide coupling agents suitable

for the uses described herein see M. Bodansky, *supra*. Normally, either the amino or carboxyl group of the first amino acid residue is protected by a suitable, selectively removable protecting group. A different protecting group is utilized for amino acids containing a reactive side group, e.g., lysine. A variety of protecting groups known in the field of peptide synthesis and recognized by conventional abbreviations therein, may be found in T. Greene, Protective Groups In Organic Synthesis, Academic Press (1981).

In order to determine the presence of an active Sema-H polypeptide, functional assays or antibody recognition of specific epitopes can be used.

Therefore, in addition to the Sema-H polypeptides, the invention also provides antibodies to Sema-H polypeptides. Such antibodies are immunoglobulin molecules or portions thereof that are immunologically reactive with a Sema-H polypeptides of the present invention. It should be understood that the antibodies of this invention include antibodies immunologically reactive with fusion proteins and multimeric proteins comprising a Sema-H polypeptide. The generation of antibodies may be achieved by standard methods in the art for producing polyclonal and monoclonal antibodies using a Sema-H polypeptide or fragment thereof as antigen. Such antibodies or active fragments thereof (such as Fab, Fab', $F(ab)_2$ fragments), may be used to assay *in vivo* or *in vitro* levels of Sema-H polypeptide. Increased *in vivo* levels of Sema-H polypeptides may be indicative of the onset of metastasis. In a preferred embodiment such antibodies would be generated using fragments of semaphorin-H that

represent unique portions of the molecule, as described previously.

An antibody of this invention may also be a hybrid molecule formed from immunoglobulin sequences from different species (e.g., mouse and human) or from portions of immunoglobulin light and heavy chain sequences from the same species. It may be a molecule that has multiple binding specificities, such as a bifunctional antibody prepared by any one of a number of techniques known to those of skill in the art including: the production of hybrid hybridomas; disulfide exchange; chemical cross-linking; addition of peptide linkers between two monoclonal antibodies; the introduction of two sets of immunoglobulin heavy and light chains into a particular cell line; and so forth.

The antibodies of this invention may also be human monoclonal antibodies produced by any of the several methods known in the art. For example, human monoclonal antibodies may be produced by immortalized human cells, by SCID-hu mice or other non-human animals capable of producing "human" antibodies, by the expression of cloned human immunoglobulin genes, by phage-display, or by any other method known in the art.

Alternatively, the genetic material of antibody producing cells can be manipulated using techniques known to those of skill in the art. For example, immunoglobulin loci recovered from antibody producing cells can be manipulated to exchange the constant region for that of a different isotype or that of a human antibody, or eliminated altogether. The variable regions can be linked to encode single chain F_v regions. Multiple F_v regions can be linked to confer binding ability to more than one target or chimeric heavy and light chain combinations can be employed.

Once the genetic material is available, design of analogs which retain their ability to bind the desired target, as well as their human characteristics, is straightforward.

Once the appropriate genetic material is obtained and, if desired, modified to encode an analog, the coding sequences including those that encode, at a minimum, the variable regions of the heavy and light chain can be inserted into expression systems contained on vectors which can be transfected into standard recombinant host cells. As described herein, a variety of such host cells can be used, however mammalian cells are preferred. Typical mammalian cell lines useful for this purpose include WHO cells, COS cells and 293 cells.

The production for the antibody is then undertaken by culturing the modified recombinant host under culture conditions appropriate for the growth of the host cells and the expression of the coding sequences. The antibodies are then recovered from the culture. The expression systems are preferably designed to include signal peptides so that the resulting antibodies are secreted into the medium; however, intracellular production is also possible.

In addition to the deliberate design of modified forms of the immunoglobulin genes to produce analogs, advantage can be taken of phage display techniques to provide libraries containing a repertoire of antibodies with varying affinities for the desired antigen. For production of such repertoires, it is unnecessary to immortalize the B cells from the immunized animal; rather the primary B cells can be used directly as a source of DNA. The mixture of cDNAs obtained from B cells , e.g., derived from spleens, is

used to prepare an expression library, for example, a phage display library transfected into E. coli. The resulting cells are tested for immunoreactivity to the desired antigen. Techniques for the identification of high affinity human antibodies from such libraries are described by Griffiths, A.D. et al., EMBO J (1994) 13: 3245-3260; by Nissim, A., et al.,ibid, 692-698, and by Griffiths, A.D., et al., ibid, 725-734. Ultimately, clones from the library are identified which produce binding affinities of a desired magnitude for the antigen, and the DNA encoding the product responsible for such binding is recovered and manipulated for standard recombinant expression. Phage display libraries may also be constructed using previously manipulated nucleotide sequences and screened in a similar fashion. In general, the cDNAs encoding heavy and light chain are independently supplied or are linked to form F_v analogs for production in the phage library.

The phage library is thus screened for the antibodies with the highest affinity for the antigen and the genetic material recovered from the appropriate clone. Further rounds of screening can increase the affinity of the original antibody isolated. The manipulations described above for recombinant production of the antibody or modification to form a desired analog can then be employed.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody

molecule, or to alter it in any other way that may render it more suitable for a particular application.

Such antibodies may be monoclonal or polyclonal. Additionally, it is within the scope of this invention to include second antibodies (monoclonal or polyclonal) directed to the anti-Sema-H antibodies. The present invention further contemplates use of these antibodies in a detection assay (immunoassay) for the Sema-H gene product.

The present invention further contemplates antibodies directed against mammalian, including rat, mouse and human Sema-H proteins or polypeptides. The antibodies of the invention may be generated by using the entire Sema-H protein as an antigen or by using short peptides, encoding portions of the Sema-H protein, as antigens. When peptides are contemplated they have at least about 4 amino acids and preferably at least about 10 amino acids.

Preferably, specific peptides encoding unique portions of the mammalian Sema-H gene are synthesized for use as antigens for obtaining Sema-H antibodies.

Polyclonal antibodies directed against the Sema-H protein are prepared by injection of a suitable laboratory animal with an effective amount of the peptide or antigenic component, collecting serum from the animal, and isolating specific sera by any of the known immunoadsorbent techniques. Animals which can readily be used for producing polyclonal anti-Sema-H antibodies include chickens, mice, rabbits, rate, goats, horses and the like. Chickens are preferred because a better immune response can be obtained and because antibodies can be collected from eggs rather than by bleeding. Although the polyclonal antibodies produced by this method are utilizable in virtually any

type of immunoassay, they are generally less favored because of the potential heterogeneity of the product.

The use of monoclonal antibodies in the diagnostic or detection assays of the present invention is particularly preferred because large quantities of antibodies, all of similar reactivity, may be produced. The preparation of hybridoma cell lines for monoclonal antibody production is done by fusing an immortal cell line and the antibody producing lymphocytes. This can be done by techniques which are well known to those who are skilled in the art. (See, for example, Harlow, E. and Lane, D. Antibodies: A Laboratory Manual, Cold Spring Harbor Press, 1988; or Douillard, J.Y. and Hoffman, T., "Basic Facts About Hybridomas", in Compendium of Immunology Vol. II, L. Schwartz (Ed.), 1981.

Unlike the preparation of polyclonal sera, the choice of animal for monoclonal antibody preparation is dependent on the availability of appropriate immortal cell lines capable of fusing with the monoclonal antibody producing lymphocytes derived from the immunized animal. Mouse and rat have been the animals of choice for hybridoma technology and are preferably used. Humans can also be utilized as sources for antibody producing lymphocytes if appropriate immortalized human (or nonhuman) cell lines are available. For the purpose of making the monoclonal antibodies of the present invention, the animal of choice may be injected with from about 0.01 mg to about 20 mg of the purified Sema-H antigen. Usually the injecting material is emulsified in Freund's complete adjuvant. Boosting injections are generally also required. The separate immortalized cell lines obtained by cell fusion may be tested for

antibody production by testing the cell culture media for the ability to find the appropriate antigen.

Lymphocytes can be obtained by removing the spleen or lymph nodes of immunized animals in a sterile fashion. Alternately, lymphocytes can be stimulated or immunized in vitro, as described, for example, in C. Reading, J. Immunol. Meth., 53: 261-291 (1982). To immortalize the monoclonal antibody producing lymphocytes, the lymphocytes must be fused to immortalized cells. A number of cell lines suitable for fusion have been developed, and the choice of any particular line for hybridization protocols is directed by any one of a number of criteria such as speed, uniformity of growth characteristics, deficiency of its metabolism for a component of the growth medium, and potential for good fusion frequency. Intraspecies hybrids, particularly between like strains, work better than interspecies fusions.

Several cell lines are available, including mutants selected for the loss of ability to create myeloma immunoglobulin. Included among these are the following mouse myeloma lines: MPC-X45-6TG, P3 NS1/1-Ag4-1, P3-X63-Ag14 (all BALB/C derived), Y3'Ag1.2.3 (rat), and U266 (human).

Cell fusion can be induced either by virus, such as Epstein-Barr or Sendai virus, or polyethylene glycol. Polyethylene glycol (PEG) is the most efficacious agent for the fusion of mammalian somatic cells. PEG itself may be toxic for cells, and various concentrations should be tested for effects on viability before attempting fusion. The molecular weight range of PEG may be varied from 1,000 to 6,000. It give best results when diluted to from about 20% to about 70% w/w in saline or serum-free medium. Exposure

to PEG at 37°C for about 30 seconds is preferred in the present case, utilizing murine cells. Extremes of temperature (i.e. about 45°C) are avoided, and preincubation of each component of the fusion system at 37°C prior to fusion gives optimum results. The ratio between lymphocytes and immortalized cells optimized to avoid cell fusion amongst lymphocytes ranges of from about 1:1 to about 1:10.

The successfully fused cells can be separated from the immortalized cell line by any technique known by the art. The most common and preferred method is to choose an immortalized cell line which is Hypoxanthine Guanine Phosphoribosyl Transferase (HGPRT) deficient. Since these cells will not grow in an aminopterin-containing medium, only hybrids of lymphocytes and immortalized cells will grow. The aminopterin containing medium is generally composed of hypoxanthine 1×10^{-4}M, aminopterin 1×10^{5}M, and thymidine 3×10^{-5}M, commonly known as the HAT medium. Fused cells are generally grown for two weeks and then fed with either regular culture medium or hypoxanthine, thymidine- containing medium.

The fused cell colonies are then tested for the presence of antibodies that recognize the Sema-H protein. Detection of hybridoma antibodies can be performed using an assay where the antigen is bound to a solid support and allowed to react to hybridoma supernatants containing putative antibodies. The presence of antibodies may be detected by "sandwich" techniques using a variety of indicators. Most of the common methods are sufficiently sensitive for use in the range of antibody concentrations secreted during hybrid growth.

Cloning of hybrid cells can be carried out after 20-25 days of cell growth in selected medium. Cloning can be performed by cell limiting dilution in fluid phase or by directly selecting single cells growing in semi-solid agarose. For limiting dilution, cell suspensions are diluted serially to yield a statistical probability of having only one cell per well. For the agarose techniques, hybrids are seeded in a semisolid upper layer, over a lower layer containing feeder cells. The colonies from the upper layer may be picked up and eventually transferred to wells.

Antibody-secreting hybrid cells can be grown in various tissue culture flasks, yielding supernatants with variable concentrations of antibodies. In order to obtain higher concentrations, hybrid cells may be transferred into animals to obtain inflammatory ascites. Antibody- containing ascites can be harvested 8-12 days after intraperitoneal injection. The ascites contain a higher concentration of antibodies but include both monoclonals and immunoglobulins from the inflammatory ascites. Antibody purification may then be achieved by, for example, affinity chromatography.

Finally, Sema-H polypeptide antibodies may be employed, for example, to purify Sema-H polypeptides from either native sources or transformed hosts expressing Sema-H polypeptides. For example, the antibodies can be conjugated by standard techniques to an insoluble matrix to form an immunoaffinity resin.

The present invention further provides methods for diagnosing metastatic cancer and for distinguishing metastatic tumors from benign tumors. The data presented herein demonstrates that Sema-H expression is higher in metastatic tumor cells than in

normal or non-metastatic tissue. Such metastatic tumors could be derived from the lung, liver, kidney, mammary gland, epithelial, thyroid, leukemic, pancreatic, endometrial, ovarian, cervical, skin, colon or lymphoid tissue and compared to benign tumor cells or the corresponding normal cells. Accordingly, in one aspect of the invention, metastatic cancer can be detected in patient's serum by a simple immunoassay. Moreover, metastatic cancer can also be diagnosed in tissue biopsies by the present immunoassays or by in situ hybridization assays.

In accordance with the present inventive discovery, the increased expression of the Sema-H gene in a cell or tissue is strongly indicative of metastatic potential. The present invention utilizes this unexpected and surprising correlation of high mammalian Sema-H gene expression with high metastatic potential to detect or diagnose malignant cancer. Both the mammalian Sema-H nucleic acid and antibodies directed against mammalian Sema-H proteins are contemplated for the diagnosis of malignant cancer.

A nucleic acid probe of the present invention may be any portion or region of a mammalian Sema-H RNA or DNA sufficient to give a detectable signal when hybridized to Sema-H mRNA derived from a tissue sample. The nucleic acid probe produces a detectable signal because it is labeled in some way, for example because the probe was made by incorporation of nucleotides linked to a "reporter molecule".

A "reporter molecule", as used herein, means any molecule which, by its chemical nature, provides an analytically identifiable signal allowing detection of the hybridized probe. Detection may be either quantitative or non-quantitative. One of skill in the

art will recognize that although any reporter molecule known in the art can be used, the most commonly used reporter molecules in this type of assay are enzymes, fluorophores and radionuclides. Such reporter molecules typically are covalently linked to nucleotides which are incorporated into a Sema-H DNA or RNA. Examples of commonly used enzymes include horseradish peroxidase, alkaline phosphatase, glucose oxidase and β-galactosidase. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the corresponding enzyme, of a detectable color change. For example, p-nitrophenyl phosphate is suitable for use with alkaline phosphatase conjugates; for horseradish peroxidase, 1,2-phenylenediamine, 5-aminonalicyclic acid or tolidine are commonly used.

Examples of fluorescent labeling agents are fluorescein isocyanate (FIC), fluorescein isothiocyanate (FITC), and the like. Examples of radioactive element are ^{125}I or ^{51}Cr which produce gamma ray emissions, or a radioactive element that emits positrons which produce gamma rays upon encounters with electrons present in the test solution, such as ^{11}C, ^{15}O, or ^{13}N. Detection may also be by other methods, for example via avidin-biotin complexes.

Incorporation into a Sema-H probe may be by nick translation, random oligo priming, by 3' or 5' end labeling, by labeled single-stranded DNA probes using single-stranded bacteriophage vectors (e.g. M13 and related phage), or by other means known in the art, (Sambrook et al., 1989, Molecular Cloning, A laboratory Manual. Cold Spring Harbor Laboratory Press. Pages 10.1-10.70). Incorporation of a reporter molecule into a Sema-H RNA probe may be by synthesis of Sema-H RNA

using T3, T7, Sp6 or other RNA polymerases (Sambrook et al., supra: 10.27-10.37).

Detection or diagnosis of metastatic cancer by the nucleic acid probe of the present invention can be by a variety of hybridization-techniques which are well known in the art. In one embodiment, patient tissue specimens are sectioned and placed onto a standard microscope slide, then fixed with an appropriate fixative. The Sema-H RNA or DNA probe, labeled by one of the techniques described above, is added. The slide is then incubated at a suitable hybridization temperature (generally 37°C to 55°C) for 1-20 hours. Non-hybridized RNA or DNA probe is then removed by extensive, gentle washing. If a non-radioactive reporter molecule is employed in the probe, the suitable substrate is applied and the slide incubated at an appropriate temperature for a time appropriate to allow a detectable color signal to appear as the slide is visualized under light microscopy. Alternatively, if the Sema-H probe is labeled radioactively, slides may be dipped in photoemulsion after hybridization and washing, and the signal detected under light microscopy after several days, as exposed silver grains.

In another embodiment, metastatic cancer is detected from RNA derived from tissue specimens by the Sema-H nucleic acid probe. RNA is isolated using methods known in the art and from specimens fixed onto nitrocellulose or nylon filters. Specimen mRNA can be purified, or specimen cells may be simply lysed and cellular mRNA fixed unto a filter. Specimen mRNA can be size fractionated through a gel before fixation onto a filter, or simply dot blotted unto a filter. Sema-H gene expression is then detected using probes of the

invention and hybridization techniques that are well known in the art.

In a further embodiment, a kit for the detection of Sema-H mRNA is provided. In general, a kit for detection of Sema-H mRNA contains at least one Sema-H nucleic acid of the invention. Such a Sema-H nucleic acid probe may be labeled with a reporter molecule or unlabeled. The kit may include an unlabeled Sema-H nucleic acid which can be modified by the kit user to include a reporter molecule, for example by nick translation or RNA transcription.

In a further embodiment, such a kit would be directed towards the quantitation of Sema-H mRNA in a cell using techniques known to those of skill in the art, for example, Northern Blot analysis or RT-PCR. Quantification of Sema-H mRNA based on appropriate standards, easily determined by one of skill in the art using the disclosed Sema-H DNA sequence, would distinguish metastatic and potentially metastatic cells from non-metastatic cells.

In another embodiment, the kit is compartmentalized: a first container can contain Sema-H RNA at a known concentration to act as a standard or positive control, a second container can contain Sema-H DNA suitable for synthesis of a detectable nucleic acid probe, and a third and a fourth container can contain reagents and enzymes suitable for preparing said Sema-H detectable probe. If the detectable nucleic acid probe is made by incorporation of an enzyme reporter molecule, a additional containers can contain a substrate, or substrates, for the enzyme provided.

Another embodiment of the present invention provides a method for diagnosing metastatic cancer by contacting a biological sample taken from an individual

to be tested for metastatic cancer with an antibody of the invention. Detection of an antigen-antibody complex in this immunoassay is diagnostic of metastatic cancer. According to this embodiment, the antibody of the invention can further be brought into contact with a specimen of bodily fluids, bodily secretions or tissue from an individual to be tested for the presence of the semaphorin-H antigen under conditions sufficient to form an antibody-antigen complex. The body fluids are, for example, blood and urine. The bodily secretions are, for example, urine, tears, sweat, saliva, cervical secretions, vaginal secretions, mucosal secretions or intraperitoneal ascitic fluid. The tissue can be, but does not have to be, excised from the individual.

In a preferred embodiment, the present invention provides a method for diagnosing metastatic cancer which involves contacting a sample from an individual to be tested for such cancer with an antibody reactive with a mammalian Sema-H protein or an antigenic fragment thereof, for a time and under conditions sufficient to form an antigen-antibody complex, and detecting the antigen-antibody complex.

The presence of the Sema-H-protein, or its antigenic components, in a patient's serum, tissue or biopsy sample can be detected utilizing antibodies prepared as above, either monoclonal or polyclonal, in virtually any type of immunoassay. A wide range of immunoassay techniques are available as can be seen by reference to Harlow, et al. (Antibodies: A Laboratory Manual, Cold Spring Harbor Press, 1988) and U.S. Patent No. 4,016,043 and 4,424,279. This, of course, includes both single-site and two-site, or "sandwich" of the noncompetitive types, as well as in traditional

competitive binding assays. Sandwich assays are among the most useful and commonly used assays. A number of variations of the sandwich assay technique exist, and all are intended to be encompassed by the present invention. Briefly, in a typical forward assay, an unlabeled antibody is immobilized in a solid substrate and the sample to be tested brought into contact with the bound molecule. After a suitable period of incubation, for a period of time sufficient to allow formation of an antibody-antigen binary complex, a second antibody, labeled with a reporter molecule capable of producing a detectable signal is then added and incubated, allowing time sufficient for the formation of a ternary complex of antibody-labeled antibody. Any reacted material is washing way, and the presence of the antigen is determined by observation of a signal produced by the reporter molecule. The results may either be qualitative, by simple observation of the visible signal, or may be quantitated by comparing with a control sample containing known amounts of hapten. Variations on the forward assay include a simultaneous assay, in which both sample and labeled antibody are added simultaneously to the bound antibody, or a reverse assay in which the labeled antibody and sample to be tested are first combined, incubated and then added to the unlabeled surface bound antibody. These techniques are well known to those skilled in the art, and then possibly of minor variations will be readily apparent. As used herein, "sandwich assay" is intended to encompass all variations on the basic two-site technique.

The Sema-H protein may also be detected by a competitive binding assay in which a limiting amount of

antibody specific for the Sema-H protein is combined with specified volumes of samples containing an unknown amount of the Sema-H protein and a solution containing a delectably labeled known amount of the Sema-H protein. Labeled and unlabeled molecules then compete for the available binding sites on the antibody. Phase separation of the free and antibody-bound molecules allows measurement of the amount of label present in each phase, thus indicating the amount of antigen or hapten in the sample being tested. A number of variations in this general competitive binding assays currently exist.

In any of the known immunoassays, for practical purposes, one of the antibodies or the antigen will be typically bound to a solid phase and a second molecule, either the second antibody in a sandwich assay, or, in a competitive assay, the known amount of antigen, will bear a detectable label or reporter molecule in order to allow visual detection of an antibody-antigen reaction. When two antibodies are employed, as in the sandwich assay, it is only necessary that one of the antibodies be specific for the Sema-H protein or its antigenic components. The following description will relate to a discussion of a typical forward sandwich assay; however, the general techniques are to be understood as being applicable to any of the contemplated immunoassays.

In the typical forward sandwich assay, a first antibody having specificity for the Sema-H protein or its antigenic components is either covalently or passively bound to a solid surface. The solid surface is typically glass or a polymer, the most commonly used polymers being cellulose, polyacrylamide, nylon, polystyrene, polyvinyl chloride or

polypropylene. The solid supports may be in the form of tubes, beads, discs or microplates, or any other surface suitable for conducting an immunoassay. The binding processes are well-known in the art and generally consisting of crosslinking covalently binding or physically adsorbing the molecule to the insoluble carrier. Following binding, the polymer-antibody complex is washed in preparation for the test sample. An aliquot of the sample to be tested in then added to the solid phase complex and incubated at a suitable temperature ranging from about 4°C to about 37°C (for example 25°C) for a period of time sufficient to allow binding of any subunit present in the antibody. The incubation period will vary but will generally be in the range of about 2-40 minutes to several hours. Following the incubation period, the antibody subunit solid phase is washed and dried and incubated with a second antibody specific for a portion of a Sema-H hapten. The second antibody is linked to a reporter molecule which-is used to indicate the binding of the second antibody to the hapten.

The most commonly used reporter molecules in this type of assay are either enzymes, fluorophores or radionuclide containing molecules. In the case of an enzyme immunoassay, an enzyme is conjugated to the second antibody, generally by means of glutaraldehyde or periodate. As will be readily recognized, however, a wide variety of different conjugation techniques exist, which are readily available to the skilled artisan. Commonly used enzymes include horseradish peroxidase, glucose oxidase, β-galactosidase and alkaline phosphates, among others. The substrates to be used with the specific enzymes are generally chosen for the production, upon hydrolysis by the

corresponding enzyme, of a detectable color change. For example, p-nitrophenyl phosphate is suitable for use with alkaline phosphatase conjugates; for peroxidase conjugates, 1,2-phenylenediamine, 5-aminosalicyclic acid, or tolidine are commonly used. It is also possible to employ fluorogenic substrates, which yield 8 fluorescent product rather than the chromogenic substrates noted above. In all cases, the enzyme-labeled antibody is added to the first antibody hapten complex, allowed to bind, and then the excess reagent is washed away. A solution containing the appropriate substrate is then added to the ternary complex of antibody-antigen-antibody. The substrate will react with the enzyme linked-to the second antibody, giving a qualitative visual signal, which may be further quantitated, usually spectrophotometrically, to give an indication of the amount of hapten which was present in the sample.

Alternately, fluorescent compounds, such as fluorescein and rhodamine, may be chemically coupled to antibodies without altering their binding capacity. When activated by illumination with light of a particular wavelength, the fluorochrome-labeled antibody absorbs the light energy, inducing a state of excitability in the molecule, followed by emission of the light at a characteristic color visually detectable with a light microscope. The fluorescent labeled antibody is allowed to bind to the first antibody-hapten complex. After washing off the unbound reagent, the remaining ternary complex is then exposed to the light of the appropriate wavelength, the fluorescence observed indicates the presence of the hapten of interest. Immunofluorescence techniques are very well established in the art. However, other reporter

molecules, such as radioisotope, chemiluminescent or bioluminescent molecules, may also be employed. It will be readily apparent to the skilled technician how to vary the procedure to suit the required purpose.

In another embodiment, the antibodies directed against the Sema-H protein may be incorporated into a kit for the detection of the Sema-H protein. Such a kit may encompass any of the detection systems contemplated and described herein, and may employ either polyclonal or monoclonal antibodies directed against the Sema-H protein. Both Sema-H antibodies complexed to a solid surface described above or soluble Sema-H antibodies are contemplated for use in a detection kit. A kit of the present invention has at least one container having an antibody reactive with a mammalian Sema-H polypeptide. However, the present kits can have other components. For example, the kit can be compartmentalized: the first container contains Sema-H protein as a solution, or bound to a solid surface, to act as a standard or positive control, the second container contains anti-Sema-H primary antibodies either free in solution or bound to a solid surface, a third container contains a solution of secondary antibodies covalently bound to a reporter molecule which are reactive against either the primary antibodies or against a portion of the Sema-H protein not reactive with the primary antibody. A fourth and fifth container contains a substrate, or reagent, appropriate for visualization of the reporter molecule. The kit of this embodiment could be further designed to perform techniques known to those of skill in the art, such as immunoblotting, dot blotting and Western assays using the antibodies of the present invention.

Quantification of Sema-H polypeptide based upon appropriate standards known to those of skill in the art is used for distinguishing metastatic and potentially metastatic cells from non-metastatic cells.

The subject invention therefore encompasses polyclonal and monoclonal antibodies useful for the detection of Sema-H protein as a means of diagnosing metastatic cancer. Said antibodies may be prepared as described above, then purified, and the detection systems contemplated and described herein employed to implement the subject invention.

The present invention also contemplates treating metastatic cancers and tumors by inactivating, destroying or nullifying the Sema-H gene or protein, or cells expressing the Sema-H gene. The treatment of cancer, as described in the specification and claims, contemplates preferably lung, liver, kidney, thyroid, mammary gland, leukemic, pancreatic, endometrial, ovarian, cervical, skin, colon or lymphoid cancers.

A preferred use then, according to this embodiment, is in the treatment of cancer in a patient. The term "patient", as used herein refers to any mammal, especially humans. DNA sequences encoding inactive Sema-H polypeptides would be useful in gene therapy to inhibit the metastatic potential of the cell. According to this embodiment, the inactive Sema-H polypeptide encoded by this DNA would compete for substrate with native, active Sema-H. This would decrease the Sema-H activity in the cell, inhibiting metastasis. Preferred cells to be treated by this method are cancer cells.

In another embodiment, antibodies, prepared as described above, may be utilized to direct toxic events to Sema-H protein expressing cells. Either

unconjugated anti-sema-H antibodies or anti-Sema-H antibodies conjugated to a therapeutic molecule may be employed in the therapy of cancer. For example, it may be advantageous to couple the antibodies of this invention to toxins such as diphtheria, pseudomonas exotoxin, ricin A chain, gelonin, etc., or antibiotics such as penicillins, tetracyclines and chloramphenicol.

Moreover, the present invention provides a method of inhibiting metastasis in a cancer cell by introducing into the cancerous cell an anti-sense nucleic acid sequence. Antisense Sema-H nucleic acids can inhibit metastatic cancer by binding to sense Sema-H mRNA. Such binding can either prevent translation of Sema-H protein or destroy Sema-H sense mRNA, e.g., through the action of RNaseH. Accordingly, less Sema-H protein is available to potential metastatic tumor cells and metastasis of these cells is inhibited. In one embodiment, such an antisense nucleic acid is a DNA or RNA molecule having at least 10 nucleotides of the antisense strand of [SEQ ID NO:1]. Preferably, the antisense Sema-H nucleic acids of the present invention have at least 15 or 17 nucleotides.

In one embodiment, this method employs an expression vector including a nucleic acid encoding an antisense nucleotide sequence for Sema-H operably linked to a segment of the vector which can effect expression of an antisense Sema-H RNA. Any of the foregoing expression vectors which can express high levels of Sema-H RNA can be used for this method.

Another embodiment of the present invention provides pharmaceutical compositions comprising an antibody reactive with a mammalian Sema-H polypeptide or fragment or derivative thereof, an antisense Sema-H nucleic acid or a Sema-H protein.

The active ingredients of a pharmaceutical composition containing the Sema-H protein (i.e. pro-metastatic reagent) or anti-Sema-H antibodies and antisense Sema-H nucleic acids (i.e. anticancer reagents) are contemplated to exhibit effective therapeutic activity, for example, in treating cancer. Thus the active ingredients of the therapeutic compositions containing Sema-H protein anti-cancer reagents, are administered in therapeutic amounts which depend on the particular disease, route of administration, and other factors well known to those of skill in the art. Preferably, from about 0.5 µg to about 2000 mg per kilogram of body weight per day may be administered. The dosage regimen may be adjusted to provide the optimum therapeutic response. For example, several divided doses may be administered daily or the dose may be proportionally reduced as indicated by the exigencies of the therapeutic situation. The active compound may be administered in a convenient manner such as by the oral, intravenous (where water soluble), intramuscular, subcutaneous, intranasal, intradermal or suppository routes. Depending on the route of administration, the active ingredients may be required to be coated in a material to protect said ingredients from the action of enzymes, acids and other natural conditions which may inactivate said ingredients. In order to administer Sema-H protein or anti-cancer reagents of this invention by other than parenteral administration, they should preferably be coated by, or administered with, a material to prevent its inactivation. For example, Sema-H protein or anti-cancer reagents may be administered in an adjuvant, co-administered with enzyme inhibitors or in liposomes. Adjuvants useful for this purpose include resorcinols,

nonionic surfactants such as polyoxyethylene oleyl ether and n-hexadecyl polyethylene ether. Enzyme inhibitors include pancreatic trypoin inhibitor, diisopropylfluorophosphate (DFP) and trasylol. Liposomes include water-in-oil-in-water P40 emulsions as well as conventional liposomes.

The anti-Sema-H proteins, antibodies, or anti-Sema-H nucleic acids may also be administered parenterally or intraperitoneally. Dispersions can also be prepared in glycerol, liquid polyethylene glycols, and mixtures thereof, and in oils. Under ordinary conditions of storage and use, these preparations contain a preservative to prevent the growth of microorganisms.

The pharmaceutical forms suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. In all cases the form must be sterile and must be fluid to the extent that easy syringe ability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, pall (for example, glycerol, propylene glycol, liquid polyethylene glycol, and the like), suitable mixtures thereof and vegetable oils. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. The preventions of the action of microorganisms can be brought about by various antibacterial and antifungal

agents, for example, parabens, chlorobutanol, phenol, sorbic acid, thimerosal, and the like. In many cases it will be preferable to include isotonic agents, for example, sugars or sodium chloride. Prolonged absorption of the injectable compositions can be brought about by the use in the compositions of agents delaying absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions are prepared by incorporating the active compounds in the required amount in the appropriate solvent with various of the other ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the various sterilized active ingredients into a sterile vehicle which contains the basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum-drying and the freeze-drying technique which yield a powder of the active ingredient plus any additional desired ingredient from previously sterile-filtered solution thereof.

When the Sema-H protein or anti-cancer reagents are suitably protected as described above, the active compound may be orally administered, for example, with an inert diluent or with an assimilable edible carrier, or it may be enclosed in hard or soft shell gelatin capsule, or it may be compressed into tablets, or it may be incorporated directly with the food of the diet. For oral therapeutic administration, the active compound may be incorporated with excipients and used in the form of ingestible tablets, buccal tablets, trochee, capsules, elixirs, suspensions,

syrups, wafers, and the like. Compositions or preparations according to the present invention are prepared so that an oral dosage unit form contains between about 0.5 µg and 2000 µg of active compound.

The tablets, troches, pills, capsules, and the like, as described above, may also contain the following: a binder such as gum tragacanth, acacia corn starch or gelatin; excipients such as dicalcitim phosphate; a disintegrating agent such as corn starch, potato starch, alginic acid, and the like; a lubricant such as magnesium stearate; and a sweetening agent such as sucrose, lactose or saccharin may be added or a flavoring agent such as peppermint, oil or wintergreen or cherry flavoring. When the dosage unit form is a capsule it may contain, in addition to materials of the above type, a liquid carrier. Various other materials may be present as coatings or to otherwise modify the physical form of the dosage unit. For instance, tablets, pills or capsules may be coated with shellac, sugar or both. A syrup or elixir may contain the active compound, sucrose as a sweetening agent, methyl and propylparabens as preservatives, a dye and flavoring such as cherry or orange flavor. Of course, any material used in preparing any dosage unit form should be pharmaceutically pure and substantially non-toxic in the amounts employed. In addition, the active compound may be incorporated into sustained-release preparations and formulations.

It is especially advantageous to formulate parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the mammalian subjects to be treated; each unit containing a predetermined

quantity of the active material calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the novel dosage unit forms of the invention are dictated by and directly depending on (a) the unique characteristics of the active material and the particular therapeutic effect to be achieved, and (b) the limitations inherent in the art of compounding such as active material for the treatment of disease in living subjects having a diseased condition in which bodily health is impaired as herein disclosed in detail.

The principal active ingredient is compounded for convenient and effective administration in effective amounts with a suitable pharmaceutically acceptable carrier in dosage unit form as hereinbefore disclosed. A unit dosage form can, for example, contain the principal active compound in amounts ranging from 0.5 µg to about 2000 µg. Expressed in proportions, the active compound is generally present in from about 10 µg to about 2000 mg/ml of carrier. In the case of compositions containing supplementary active ingredients, the dosages are determined by reference to the usual dose and manner of administration of the said ingredients.

As used herein "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and adsorption delaying agents, and the like. The use of such media gents for pharmaceutical active substances is well known in the art. Except insofar as any conventional media or agent in incompatible with the active ingredient, use thereof in the therapeutic compositions is contemplated. Supplementary active

ingredients can also be incorporated into the compositions.

The Sema-H DNA sequences of this invention, of fragments thereof, may be used to identify homologous DNA sequences so as to potentially identify other members of this family.

These DNA sequences, or fragments thereof, are useful in a number of therapeutic and diagnostic applications. According to one embodiment, the DNA sequences encoding a Sema-H polypeptide may be used in either *in vitro* or *in vivo* gene therapy. In this embodiment, the DNA must be contained in a suitable vehicle for gene therapy. Such vehicles are known in the art. For example, various viruses that are capable of transferring genetic material to a target cell, without replicating in that cell have been described (United States patents 5,112,767, 5,240,846 and 5,112,767, the disclosures of which are herein incorporated by references). Such viruses include replication-defective adenoviruses, adeno-associated viruses (AAV), and replication defective retroviruses, such as PLJ, pZip, pWe and pEM.

Gene delivery systems other than viruses can also be employed in the methods of this invention. For example, the gene to be transferred may be packaged in a liposome. When cells are incubated with DNA-encapsidated liposomes, they take up the DNA and express it. To form these liposomes, one mixes the DNA of an expression vector which expresses the gene to be transferred with lipid, such as N-[1-(2,3,dioleyloxy)propyl]-N,N,N-tri-methylammonium chloride (DOTMA) in a suitable buffer, such as Hepes buffered saline. This causes the spontaneous formation of lipid-DNA complexes (liposomes) which can be

employed in the methods of this invention [P. L. Felgner et al., Proc. Natl. Acad. Sci. USA, 84, pp. 7413-17 (1987)].

Another gene delivery system that may be utilized in this invention is DNA-protein complexes. The formation of these complexes is described in United States patent 5,166,320, the disclosure of which is herein incorporated by reference. Specifically, these complexes comprise the gene to be transferred (together with promoter, enhancer sequences and other DNA necessary for expression in the target cell) linked via a suitable polymer, such as polylysine, to a polypeptide ligand for a receptor or other cell surface protein. This complex is taken up by the target cells via endocytosis after the ligand binds to the cell surface receptor. The DNA is then cleaved from the rest of the complex via intracellular enzymes which cut the polymer linker. Such complexes are particularly useful in targeting specific cell types. For examples, cancer cells may be specifically targeted through the use of proteins which specifically recognize tumor cell surface markers or receptors.

Antibodies to Sema-H polypeptides also have potential therapeutic use in preventing metastasis in a patient. Such antibodies may be directed to their target cells through conjugation to cell-specific molecules. The conjugation of such antibodies and the identity of cell-specific targeting molecules is known in the art.

Another embodiment of the present invention relates to the animal tumors and tumor cell lines developed in accordance with the present invention which are useful as model systems of the metastatic process. These tumors and cell lines can be utilized

for screening anti-metastatic drugs and for developing therapeutic regimens for the treatment of malignant cancer is provided by the present invention.

The tumors or cell lines of the present invention each would have a highly predictable metastatic potential; however the metastatic potentials of related, but separate, tumors or cell lines can be very different. These tumors and cell lines are useful for the development of a variety of human cancer therapies, for several reasons. First, cancer cells all have similar properties, including, for example, unrestrained growth and lack of contact inhibition, which suggests that the process of cancer development is similar in all cancers. Second, the morphologies and biochemical properties of the tumors developed after injection of these tumor-derived cells are identical to analogous tumors in humans. Hence, potential anti-cancer therapies or drugs may effectively be screened by employing the animal model system of the current invention.

The utility of these unique tumors and cell lines is apparent to one skilled in the art. Briefly, animals are injected with tumors or tumor-derived cells which have a predictable metastatic potential. A proportion of the animals are treated with a potential anti-cancer drug or therapy. After a suitable period of time, all animals are sacrificed and the tissues of both treated and non-treated animals are examined for the development of primary and secondary (metastatic) tumors. If a therapeutic regimen is successful, the treated animals should have a much lower incidence of tumor formation.

Both mouse and rat model systems are provided by the present invention for the development of cancer

therapy. This is done by intramuscular transplantation or subcutaneous tail transplantations of the original spontaneous mammary tumor cells into syngeneic mice. Intramuscular transplantation has yielded a cell line called CSML-0 which has low metastatic potential. Solitary lung metastasis are detected in less than 10% of CSML-0 injected animals sacrificed because of a moribund condition. The highly metastatic CSML-100 cell line has been generated by selection of the metastatic phenotype through successive subcutaneous transplantations of CSML metastatic cells into the tail. The CSML-50 cell line, selected during the generation of CSML-100, has an intermediate level of metastatic potential.

A variety of rat tumors have been generated by irradiating normal Fischer 344 rat thyroid cell suspensions and then transplanting these cells into rats. Grafts of non-irradiated thyroid cells develop into morphologically and functionally normal thyroid tissue after transplantation into Fischer 344 syngeneic rats, if elevated levels of thyroid stimulating hormone are also provided. Irradiation of thyroid cell suspensions before transplantation has produced a series of rat thyroid carcinomas which are histopathologically identical to human counterparts.

The extensive variety of tumors and cell lines, and the varying metastatic potential of these tumors and cell lines, provides mouse and rat model systems amenable to carefully controlled studies directed towards the dissection of the metastatic process. Therapeutic regimens for treatment of malignant cancer can be developed by controlled studies of groups of animals injected with cells of high, low and intermediate metastatic potential. A drug, or

pharmaceutical composition suspected of having anti-metastatic potential, may be used to treat a proportion of animals from each group. The incidence of metastasis amongst the animals receiving the drug or pharmaceutical composition may be compared with the incidence amongst animals not receiving treatment. Therefore, the present invention provides an animal system for distinguishing effective anti-metastatic drugs and therapies from those that are ineffective.

In order that the invention described herein may be more fully understood, the following examples are set forth. It should be understood that these examples are for illustrative purposes only and are not to be construed as limiting this invention in any manner.

Example 1 - Cloning of Mouse Semaphorin-H and Correlation of SemaH Expression to Metastasis

A. Identification of Semaphorin Transcripts in Metastatic Cell Lines

Differential display was performed to compare the mRNA populations of two metastatic cell lines 66c14 and 4T1 with non-metastatic cell line 67NR. Total RNA of confluent cultures of the 4T1, 66c14 and 67NR cell lines was DNase-treated. The mRNA with UC as the final 3'-nucleotides were reverse transcribed using 5'T_{11}AG-3' primer and PCR was subsequently performed using the same primer in combination with a number of 10mer primers. The sequence of the 10mer primers is described in (17). Transcripts of the M-semaH gene were identified through the use of the 10mer primer 5'-GTTTTCGCAG-3' in the PCR amplifications. The protocol

for reverse transcription and PCR was as described (2), with the following modifications: Superscript reverse transcriptase (Gibco BRL) was used in the reverse transcription, and in the PCR, Thermoprime plus Taq-polymerase and corresponding buffer was used with 1.0mM as the final concentration of $MgCl_2$ (Advanced Biotech).

PCR conditions were also modified and were 94°C, 30s; 40°C, 2min; 72°C, 30s, for 5 remaining cycles, then 94°C, 30s; 42°C, 2min; 72°C, 30s for 35 cycles followed by 72°C, 5min. The remaining steps of electrophoresis, extraction of bands, reamplification and cloning were as described (2).

An amplified 525 base pair fragment was identified in the metastatic cell lines but not the non-metastatic cell lines and this fragment was extracted from a sequencing gel. The fragment was found to originate from a novel mouse semaphorin gene. Northern analysis was performed to demonstrate the differential expression of transcripts hybridizing to the 525 bp fragment in metastatic cell lines. The probe hybridized to three transcripts of approximately 7.0 kb, 4.5 kb and 4.0 kb in the expressing cell lines.

Examination of the ability of the cells to form lung metastasis in an experimental and spontaneous metastasis assay revealed a perfect correlation between the expression of M'semaH and the metastatic potential of the five cell lines in Figure 1B. Subsequent analysis of 14 additional cell lines identified a strong correlation between M-semaH transcripts and metastasis (Figure 1C).

B. Cloning of M-SemaH DNA

A 525 base pair fragment, amplified and extracted from 66cl4 or 4T1 metastatic cells after differential display analysis (as described above), was hybridized to a λ-cDNA library derived from the metastatic CSML-100 cell line. The cDNA library was constructed from the CSML-100 cell line using the ZAP-cDNA synthesis kit and ligated into Uni-ZAP XR vectors using ZAP-cDNA Gigapack II Gold Cloning Kit (Stratagene) as would be known to one of skill in the art. The cDNA fragments identified through hybridization were examined by PCR and restriction analysis, after which two fragments of 4435 base pairs and 3675 base pairs were sequenced.

The 3675 base pair fragment was found to be lacking a region of 478 base pairs localized 169 base pairs from the 3'-end (Figure 2A). Northern analysis with a probe specific for the 478 base pair (probe C, Figure 2C) detected only the 4.5 kb and 7.0 kb transcripts (compare probe B, Figure 2B with probe C, Figure 2C). The 3675 base pair cDNA therefore corresponds to the 4.0 kb transcript. To describe its variant nature we designated it M-semaH-v (SEQ ID NO:2)

5' RACE analysis was performed using the Marathon cDNA cloning kit (Clontech). First strand cDNA was synthesized using 1mM of oligo(dT) primer (from kit) and 1 µg of poly A⁺-RNA isolated from metastatic cell line 66cl14. Primers specific to M-semaH were 5'CTCTTAGGACCACTTGTTCAC-3' and 5'-GGGGTTGAGGAGGAAACA-3' (DNA Technology, Denmark). Other components and instructions as provided by the manufacturer. RACE analysis demonstrated shared 5' ends between the two transcripts. Northern analysis with a probe specific to the 5'untranslated portion of the 4435 bp fragment hybridized to both the 4.0 and

4.5kb transcripts (Figure 2D). Those of skill in the art will appreciate that DNA encoding additional mammalian sema-H can be cloned using these methods.

Example 2 - Expression of M-semaH Polypeptides

A. Expression in Prokaryotes

The mouse semaH open reading frame (ORF) excluding the sequence encoding the eukaryote signal peptide is cloned by PCR from the 4430 bp semaH cDNA, which is previously cloned in a pBluescript SK phagemid (vector obtained by Stratagene). The sense primer is 5'-AAACACGCATGCGAGCCCTCCTACGCCAG-3' encoding a *Sph1* site and the SemaH sequence PSYAR; and the antisense primer is 5'-CCCCCCGGGGGGTCAGGAGAGCAGCG-3' which encodes a *XmaI* site and is complementary to the sequence encoding the C-terminal part of SemaH. The PCR fragment is subcloned into the pCR2.1 vector using the TOPO-TA cloning kit (Invitrogene). Transformants are selected for ampicillin resistance. Plasmids are purified from selected colonies, restricted with *SphI* (Pharmacia) and *XmaI* enzymes (New England Biolabs), released fragments gel purified using the QIAQuick Gel Extraction Kit (Qiagen) and cloned into the *SphI* and *XmaI* sites of the pQE30 vector (Qiagen) next to a N-terminal His taq encoded by this vector. This will generate the pQE30/Mouse-His construct for protein expression in M15 bacteria (Qiagen), see below.

B. Expression in Mammalian Cells

To express the Mouse SemaH in mouse tumor cell lines we use the expression vector pcDNA3.1/(-),

which carries resistance to the antibiotic Zeocin (Invitrogen). The ORF of M-semaH is in one of two forms under the control of the CMV-promoter: 1) the ORF with the native UGA stop codon (pcDNA3.1/SemaH/zeo^R) and, 2) the ORF modified in the three last amino acids to allow the C-terminal fusion with the taq sequence NH_2-LVPRGSGPEQKLISEEDLNSAVDHHHHHH-COOH, comprising a Trombin cleavage site, a Myc epitope and a Histidin taq (pcDNA3.1/SemaH-TMH/zeo^R). In both cases the pBluescript SK phagemid containing the 4430 bp cDNA of the 5.5 kb semaH transcript is used as template for the PCR reactions.

The sense primer 5'-AGAGGAGGGCCCGCCGCCACCATGGCACC-3' encoding an *ApaI* site and a consensus *Kozak* sequence next to the native SemaH N-terminal sequence MAP will be used together with the antisense primer 5'-CGGCAGAGGGGGCCCTCAGGAGAGCAGCG-3' encoding an *ApaI* site complementary to the sequence encoding the native SemaH C-terminal sequence TLLS-stop. The fragment is subcloned into the pCR2.1 vector using the TOPO-TA cloning Kit (Invitrogen). Transformants are selected for Ampicillin resistance. Plasmids purified from selected colonies are restricted with *ApaI* enzyme (Pharmacia), released fragments gel purified using the QiaQuick Gel Extraction Kit (Qiagen) and cloned into the *ApaI* site of the vector pcDNA3.1(-)/zeo^R (Invitrogene). The resulting construct is transformed into TOP10F' bacteria (Invitrogene) and transformants are selected for resistance to Zeocin (Invitrogen).

The same sense primer as above is used together with the antisense primer 5'-GCCGGGGGCCCGCTCCCTCGGGGGACCAGCGTGTG-3' encoding an *ApaI*

site and being complementary to the native SemaH C-terminal sequence encoding HTL, skipping the stop codon and instead fused to the peptide VPRGCGP comprising a thrombin cleavage site. The fragment is subcloned as described above and after identical procedures cloned into the *ApaI* site of the pcDNA3.1(-)/Mye-HisB/neo^R vector (Invitrogene). The plasmid is transformed into TOP10F' bacteria (Invitrogen) and transformants are selected for ampicillin resistance. The C-terminal part of SemaH is fused this way to the vector encoded Mye epitope and His Taq. Plasmids, purified from selected colonies, are restricted with *PmeI* enzyme (New England Biolabs) cutting on either site of the entire semaH-TMH fragment, and cloned into the *PmeI* site of the pcDNA3.1(-)/zeo^R (Invitrogen) to generate a pcDNA3.1/SemaH-TMH/zeo^R construct. Transformants are obtained and scored as above. The inserted semaH fragment in the pcDNA3.1/SemaH-TMH construct will be used as template in a PCR amplification using the sense primer 5'-GGTCACTCTGCAGGCCCCTCCTACGCCAG-3', encoding a *PstI* site and the amino acid sequence PSYAR which follows the native N-terminal signal sequence of M-SemaH, and the antisense primer 5'-GCCGGGCGGCCGCCCCTCGGGGG-3' specific to the binding site of the antisense-primer used in constructing the pcDNA3.1/SemaH-TMH/zeo^R but encoding a *NotI* site in place of the *ApaI* site. The fragment is subcloned in a pCR2.1 vector as above, released by restriction with *NotI* and *PstI* enzymes (Pharmacia) and cloned into the respective sites of a pPICZ□B/zeo^R vector (Invitrogen) to generate a pPICZ□B/SemaH-TMH/zeo^R construct. In this construct, the semaH ORF is sitting next to the *S. Cerevisae* factor signal sequence under the control of

the *P. Pastoris* AOX1 gene promoter, which is inducible by methanol. In the C-terminal the semaH ORF is fused to a vector encoded peptide comprising a Mye epitope and a His tag similar to the one in pcDNA3.1/SemaH-TMH/zeo^R construct. Transformants are obtained and scored as above.

Example 3 - Cloning of the human SemaH homologue

Cloning of the human homologue of semaH is performed by PCR using degenerate primers which have been designed after an alignment of SemaH with known Semaphorins and Collapsins. The sense primer is 5'-CGGGATCCAT(H)TT(Y)TT(Y)TT(Y)AC(N)GA(R)AA-3' encoding the motif YFFFTEK specific to mouse SemaH and its putative chick homologue Collapsin 5, and the antisense primer is 5'-GCGGATCCTCCCA(N)GC(R)CA(R)TA(N)GG(R)TC-3' complementary to the sequence encoding the motif DPYCAWD common to all known semaphorins and collapsins. The primers encode *BamIII* sites to facilitate cloning. The PCR reaction is performed on cDNA from human fetal brain and lung (Clontech) using 5 initial cycles with low annealing temperature followed by 30 cycles with high annealing temperature. The primers amplify a region of approximately 900 bp. The obtained fragments are gel purified and cloned into the *BamHI* site of a pQE30 vector (Qiagen) fused to the vector encoded N-terminal His taq to generate the pQE30/Human-His construct. This facilitates expression and purification of the corresponding peptide in M15 bacteria using the QiaExpressionist system (Qiagen). Sequence analysis is performed using software of the Genetics Computer Group (Wisconsin, USA) to look for

evolutionary relationship between the obtained fragments and mouse SemaH. Expression of the Human homologue is examined by Northern hybridization to RNA from human Tumor Cell lines.

A *lambda* cDNA library is made from mRNA isolated from selected human tumor cell lines with the CAP-cDNA synthesis kit and ligated into UNI-ZAP XR vectors with the ZAP-cDNA Gigapack III Gold Cloning Kit (STRATAGENE). Commercial cDNA libraries from human fetal brain and lung are obtained (Clonetech or Invitrogene). The different libraries are screened with mouse semaH cDNA and human cDNA fragments isolated as described above. Obtained cDNA clones are subjected to sequence analysis to look for splice-variants, and the expression of these are evaluated comparatively in tumor and normal tissue by RT-PCR and northern hybridization.

The open reading frame of the human homologue is cloned into expression vectors for the purpose of transfection and protein expression analogous to mouse semaH above.

Example 4 - Expression of SemaH peptides, proteins and fusion proteins

A. *E. Coli*

M15 bacteria are transformed with the pQE30/Human-His construct and the pQE30/Mouse-His construct, and transformants selected for ampicillin resistance. Selected transformants are grown in LB medium, and the expression of the fusion proteins induced by IPTG and purified from total protein lysate through the use of a Ni^{2+} resin, following the guidelines of the QIAExpressionist system (Qiagen).

B. *Pichia Pastoris*

Mouse SemaH-TMH is expressed in the yeast *Pichia Pastoris* using the strains KM71, GS115 and X33 and guidelines of the EasySelect kit (Invitrogen). The pPICZ□B/SemaH-TMH construct is linearized using *SacI* restriction enzyme, and transformed into the yeast cells by electroporation. Cells are plated on agar plates containing yeast extract, peptone, dextrose, sorbitol, and from 100µg/ml-2000µg/ml Zeocin. Highly resistant clones are selected, cultured and induced with 1% v/v methanol, and tested for secretion of the SemaH-TMH to the media. Detection of the fusion protein is done by western hybridization using anti-mye antibodies (Dept. of Cell Cycle and Cancer, the Danish Cancer Society), anti-his antibodies (Qiagen), and SemaH specific antibodies. SemaH is purified using a Ni^{2+} resin according to the instructions of the QIAexpressionist system (Qiagen), combined with affinity purification using anti-mye antibodies coupled to CNBr activated sepharose.

Example 5 - Production of antibodies to mouse and human SemaH

The full-length SemaH protein or SemaH specific peptides are injected into rabbits and rats for the production of polyclonal antibodies as well as in mice for the production of monoclonal antibodies. Hybridomas are produced according to methods well known in the art, for example, through fusion of mouse spleen lymphocytes with NS-2 cells. Cells are cultured in conditioned media, and screened for production of immunoglobulins binding SemaH protein in ELISA assays and, when immobilized on membranes, by western

hybridization. The epitope specificity of selected hybridomas is determined by ELISA on a display of 20 amino acid peptides spanning the antigen with 5 amino acid overlap, each peptide is immobilized in a streptavidin coated ELISA well through biotinylation linkage.

Example 6 - Assay System for Metastatic Conversion

A cell line system developed by Cheryl J. Aslakson and Fred R. Miller at the Michigan Cancer Foundation (Aslakson, G.J. and Miller, F.R., "Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor," Cancer Research, 52, 1399-1405, 1992) is used. 67NR is a non-invasive cell line which is resistant to neomycin. 168FARN is a neomycin and 2,6-diaminopurine resistant cell line which is invasive, metastasizing to the lymph nodes, but not to the lungs. A non-metastatic cell line obtained in our lab, CSML-0, has been made neomycin resistant. Cells are transfected with the pcDNA3.1/semaH/zeo^R and pcDNA3.1/semaH-TMH/zeo^R constructs. Stable transfectants are injected into Balb/c (67NR and 168FARN) and A/sn (CSML-0) mice intravenously, subcutaneously, or in the mammary pad of female mice. Non-transfected cells will be injected in parallel as controls.

Following the guidelines of Aslakson et al., to track the path and extent of metastasis, we isolate cells from blood, lymph nodes, liver and lungs. Injection of 1 x 10^5 cells subcutaneously and in the mammary is followed by the sacrifice of groups of 5 mice at different time points thereafter. Upon

removal, blood is heparinized. Lungs and lymph nodes are digested by Collagenase IV and Elastase (Sigma Chemical Company, ICN Biochemicals) followed by dispersion in a blender (Tekmar Company). The liver is treated with Collagenase type I (Sigma) and Hyaluronidase (Sigma) followed by dispersion in the blender. Aliquots of tissue sample are plated in DME-10 media containing 10% fetal calf serum and containing combinations of the antibiotics neomycin (Gibco BRL), 2,6-diaminopurine (Sigma) and Zeocin (Invitrogene) appropriate to the cell line and plasmid. After 10-14 days the number of colonies per organ are calculated.

Differences between the transfected and non-transfected cells in the ability to survive selective events such as invasion, diffusion in the blood or integration in the lungs/liver indicates metastatic conversion.

A. Mouse

Following injections in the *subcutis* and in the mammary pad of female mice with 1 x 10^5 cells from non-metastatic, metastatic or semaH transfected cell lines, groups of 5 mice are sacrificed at different time points. Serum is collected from each sample, and the levels of SemaH protein are examined by ELISA using SemaH monoclonal antibodies. SemaH protein is quantified by calorimetric means or by radioactively labeled secondary antibodies followed by scintillation of samples. The amount of SemaH protein is correlated with parameters such as the size of the primary tumors, the pathological stage of tumor sections, immunostaining of such sections with SemaH antibodies, the presence of visible metastasis, and the presence of

clonogenic tumor cells in the blood, lymph nodes, lungs and livers obtained through the exploitation of the antibiotic resistance of the cells injected analogous to above.

B. Humans

Tissue specimens from patients with diagnosed and staged cancer are obtained. Immunohistochemistry is performed on tissue sections using the anti-SemaH monoclonal antibodies. Metastatic conversion is detected by antibody binding.

As will be appreciated by one of skill in the art, the methods and compositions disclosed herein are used for making and identifying diagnostic probes, therapeutic drugs and kits for same.

All publications and patent applications cited in this specification are herein incorporated by reference as if each individual publication or patent application were specifically and individually indicated to be incorporated by reference. Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be readily apparent to those of ordinary skill in the art in light of the teachings of this invention that certain changes and modifications may be made thereto without departing from the spirit or scope of the appended claims.

What is claimed is:

1. An isolated, recombinant or synthetic DNA molecule comprising a DNA sequence encoding a semaphorin-H polypeptide, wherein the DNA sequence is selected from:

(a) the DNA sequence of SEQ ID NO:1;

(b) the DNA sequence of SEQ ID NO: 2;

(c) the DNA sequence contained in the insert of clone ORF (EMBL No. 80941);

(d) fragments of the DNA sequences of (a-c);

(e) DNA sequences which hybridize under stringent conditions to the DNA sequences of (a-c) and which code for a polypeptide displaying Semaphorin-H activity; and

(f) DNA sequences which encode a polypeptide encoded by a DNA sequence of (a-e).

2. The isolated, recombinant or synthetic DNA molecule according to claim 1, operatively linked to an expression control sequence.

3. Host cells transformed with a DNA molecule according to claim 1 or claim 2.

4. A method for producing a polypeptide comprising the step of culturing a host cell according to claim 3.

5. Host cells expressing the polypeptide encoded by the DNA of claim 1.

6. The polypeptide produced by the method according to claim 4.

7. An isolated, recombinant or synthetic polypeptide selected from the group consisting of:

(a) the polypeptide having the amino acid sequence set forth in SEQ ID NO: 3;

(b) the polypeptide having the amino acid sequence set forth in SEQ ID NO: 4;

(c) the polypeptide encoded by the DNA sequence contained in the insert of clone SemaH (EMBL NO.93947);

(d) fragments of the foregoing polypeptides having at least 5 amino acids;

(e) a polypeptide encoded by a DNA sequence that hybridizes under stringent conditions and that is at least 80% complementary to a DNA sequence according to claim 1; and

(f) derivatives of any of the foregoing polypeptides.

8. A fusion protein comprising a polypeptide according to claim 6 or claim 7.

9. A multimeric protein comprising a polypeptide according to claim 6 or claim 7.

10. A pharmaceutical composition comprising a polypeptide according to claim 6 or claim 7; a fusion protein according to claim 8; or a multimeric protein according to claim 9.

11. An antibody that specifically binds to a polypeptide according to claim 6 or claim 7.

12. The antibody according to claim 11, wherein the antibody is polyclonal.

13. The antibody according to claim 11, wherein the antibody is monoclonal.

14. An antibody according to claim 11 or 13, wherein said antibody inhibits semaphorin activity.

15. The antibody according to claim 11, 13 or 14 wherein said antibody is labeled.

16. The antibody according to claim 15 wherein said label is selected from the group consisting of enzymes, fluorochromes, radioisotopes, and luminescers.

17. A hybridoma cell line that produces an antibody according to claim 13.

18. A method for determining the metastatic potential of a cell, comprising the step of detecting

in a biological sample the expression of a polypeptide according to claim 7.

19. The method according to claim 18, comprising the step of contacting a biological sample with an antibody according to any one of claims 11-16 and detecting the formation of an antibody-antigen complex.

20. The method according to claim 19, wherein the step of detection is by enzyme reaction, fluorescence, luminescence emission, or radioactivity measurements.

21. A method for detecting the metastatic potential of a cell, comprising the step of assaying a biological sample for the presence of RNA encoded by a DNA sequence according to claim 1.

22. The method according to claim 21, wherein the assay is RT-PCR or Northern assay.

23. A diagnostic kit comprising an antibody according to any one of claims 11-16, or a protein according to claim 6 or claim 7.

24. A diagnostic kit comprising a primer derived from a DNA sequence of claim 1.

25. A method for inhibiting metastasis, comprising the step of inhibiting the biological activity of a polypeptide according to claim 6 or claim 7.

26. The method according to claim 25, comprising the step of contacting the cells with a molecule selected from the group consisting of: an antibody according to any one of claims 11-16; a small molecule inhibitor; and a semaphorin-H ligand or fragment thereof.

27. A method for inhibiting metastasis of cells expressing a DNA sequence according to claim 1, comprising the step of preventing the expression of said DNA sequence.

28. The method according to claim 27, wherein the method utilizes an antisense nucleic acid derived from a DNA sequence according to claim 1.

29. A composition comprising an antibody according to claim 11-16 and a pharmaceutically acceptable carrier.

30. An anti-sense oligonucleotide derived from a DNA sequence according to claim 1.

31. A pharmaceutical composition comprising the antisense oligonucleotide according to claim 30.

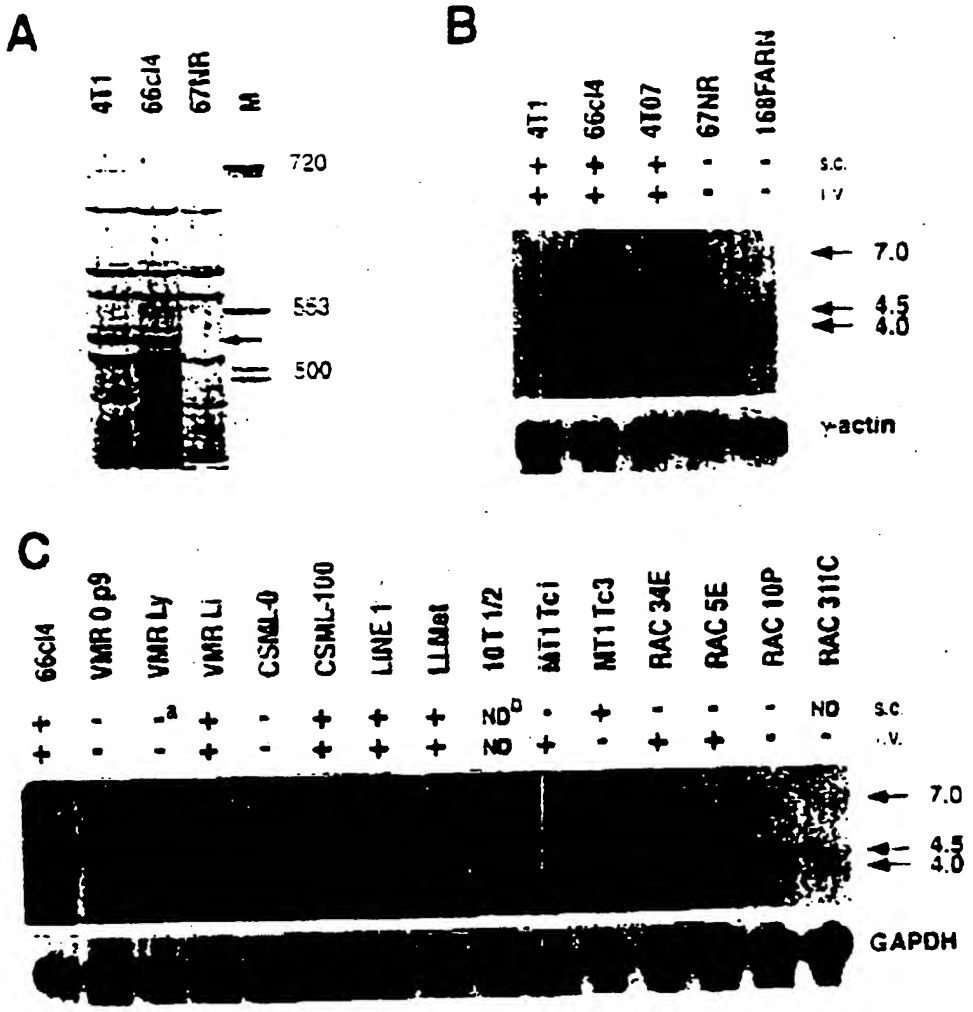
A
4T1
66cl4
67NR
M
720
553
500
B
4T1
66cl4
4T07
67NR
168FARN
s.c.
i.v.
7.0
4.5
4.0
γ-actin
C
66cl4
VMR 0 p9
VMR Ly
VMR Li
CSML-0
CSML-100
LINE 1
LLMet
10 T 1/2
MT1 Tc1
MT1 Tc3
RAC 34E
RAC 5E
RAC 10P
RAC 311C
ND
s.c.
i.v.
7.0
4.5
4.0
GAPDH

FIGURE 1

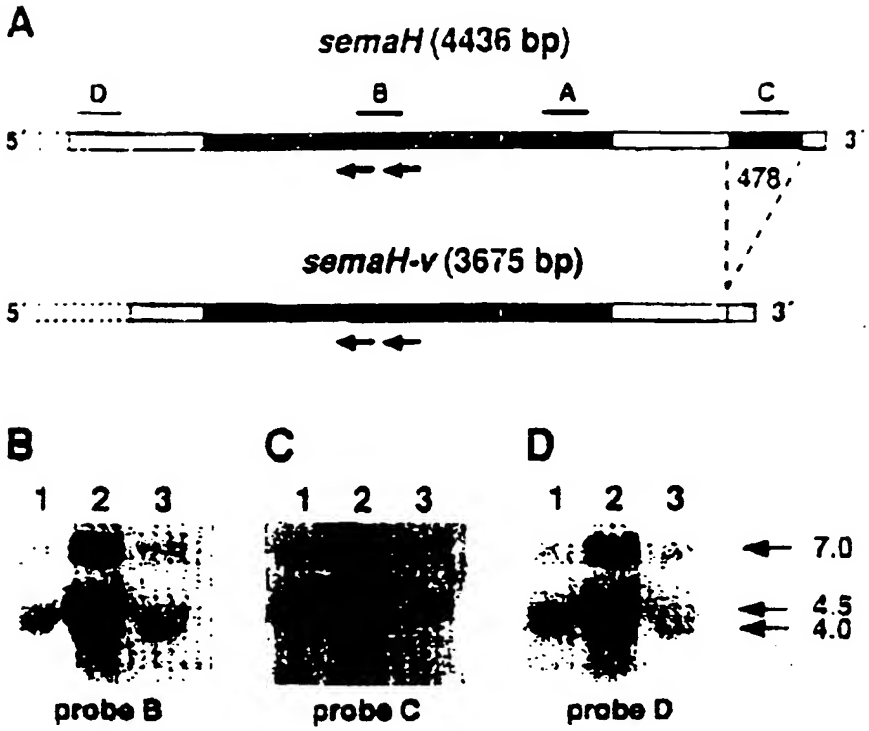
A
semaH (4436 bp)
D
B
A
C
5'
3'
478
semaH-v (3675 bp)
5'
3'
B
C
D
1 2 3
1 2 3
1 2 3
7.0
4.5
4.0
probe B
probe C
probe D

FIGURE 2

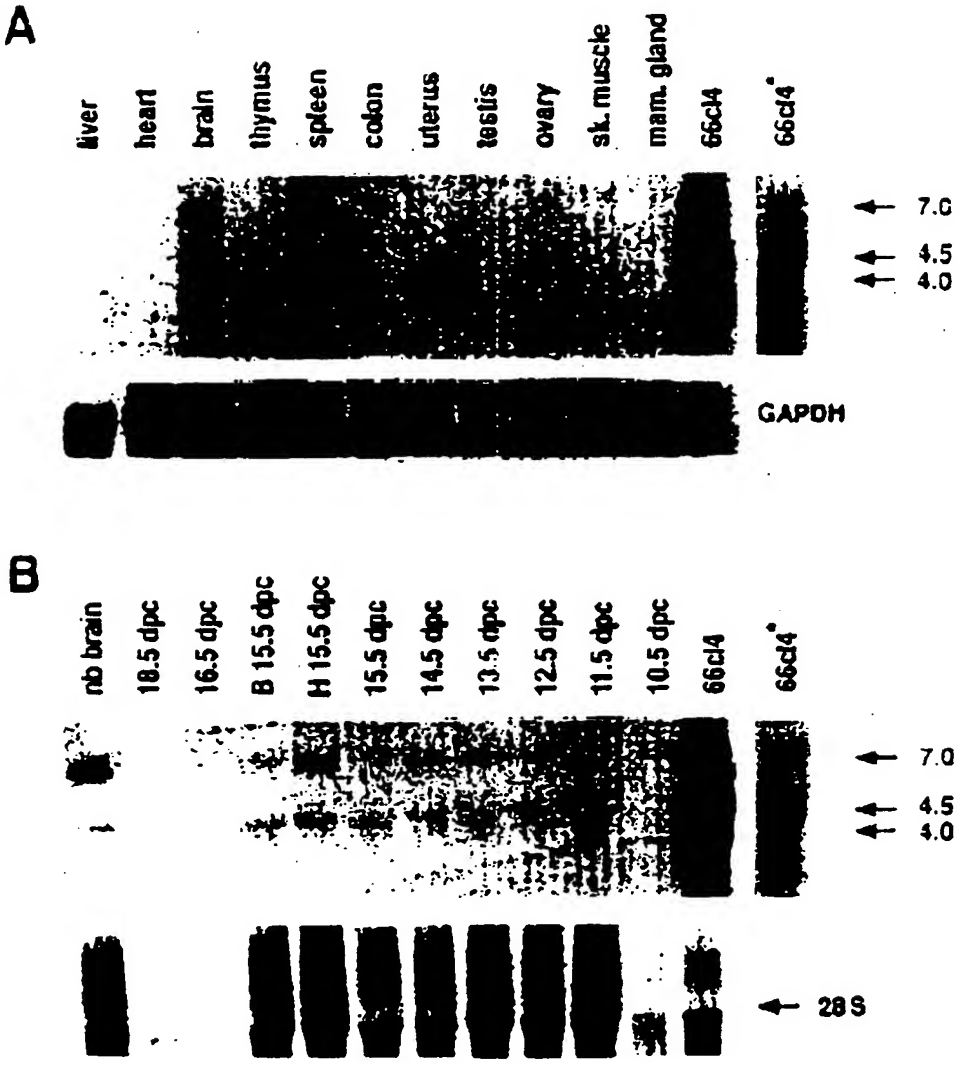
A
liver
heart
brain
thymus
spleen
colon
uterus
testis
ovary
sk. muscle
mam. gland
66cl4
66cl4*
7.0
4.5
4.0
GAPDH
B
nb brain
18.5 dpc
16.5 dpc
B 15.5 dpc
H 15.5 dpc
15.5 dpc
14.5 dpc
13.5 dpc
12.5 dpc
11.5 dpc
10.5 dpc
66cl4
66cl4*
7.0
4.5
4.0
28S

FIGURE 3

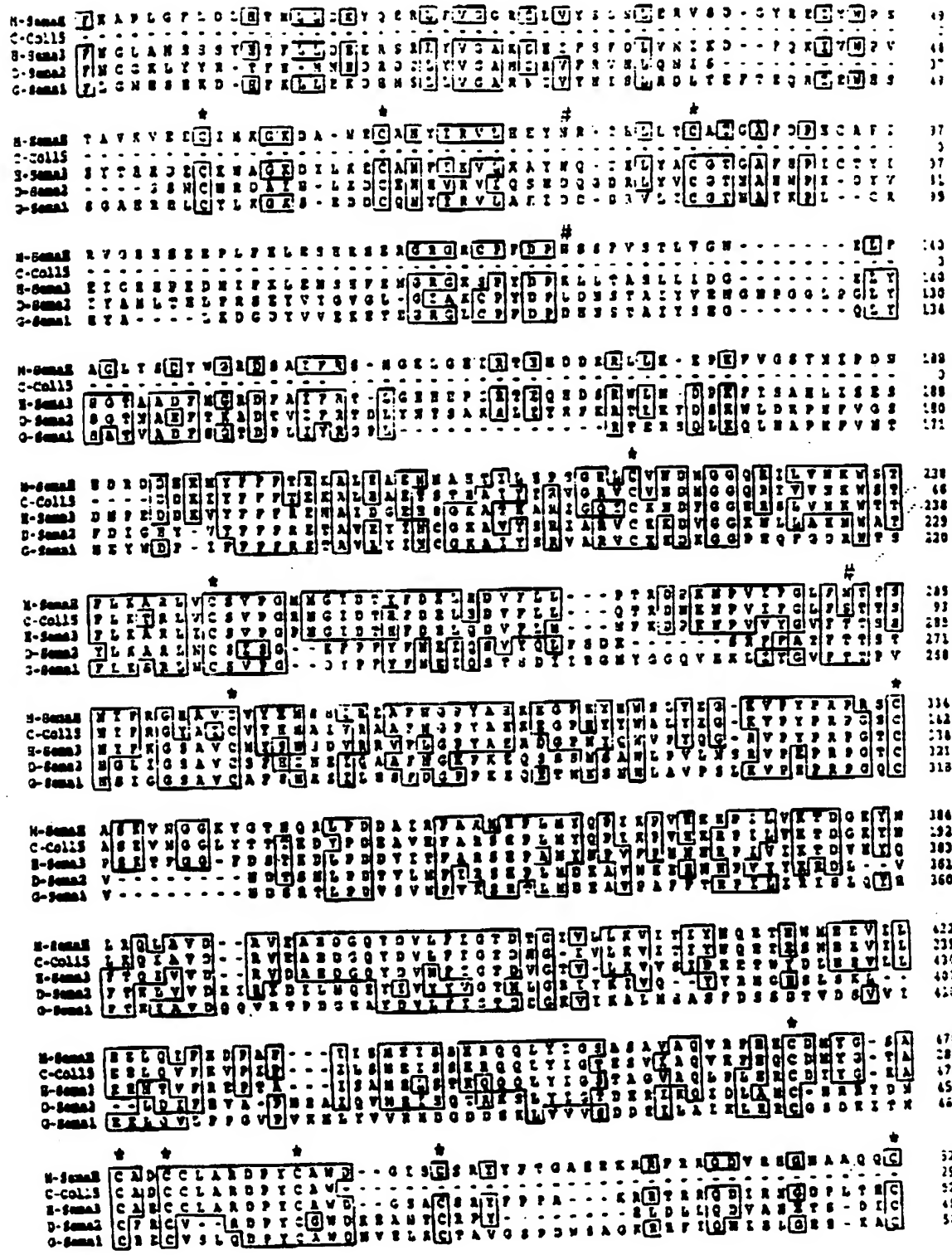

FIGURE 4A

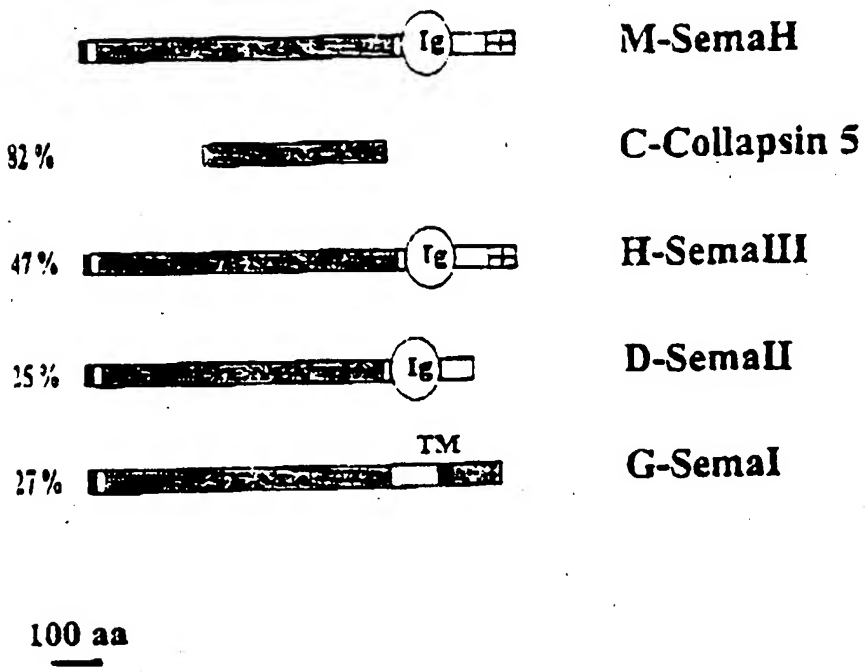

FIGURE 4B

SEQUENCES

SEQ ID NO. 1 SemaH DNA sequence

1 AATTCGGCAC GAGTTCCTAG AAACGCGCGC GGGCTCAACC CTGCCTGAAC

51 TTTCCTGTAA ACAGAGCGCT GACAGGCGGC ATCCCCGCTG GGTGGATCCC

101 GCGCCCTGGC GCAAGTGGCA CTTCTTGCTT CTAATTATCG AGAGGAGAGG

151 CGAATACGAA CTAGCTGCTC GGCAAGTCAG TGTCAGGAGG CTGACTTCTG

201 GGAGGCTGGC GGGGAGGCTG GGGGAAGAGC TGGGGGAGGC TGCTGCTCTG

251 CTTCGACTGT TTTCTCAATG AATAGCTGGC GGGGAGACTG AAGCTAGCCA

301 CAGCCTCCTC CTTCACTCCG CGTCTGGGCT GACGGCGACA GCACCAGCCC

351 GGACCTGGCT CTCAAGACGC GCTCCTTGGA CGGTCTCTTG CTCCGCGCTT

401 CTAACCACCG GGCCCAAAGA CAGAAAGGCT TAGCGGATCC AAATATTGCC

451 CGGCAAATGG CACTTGGGAA TGGTATTTTC TGATGACAAC CCCTTCTGTT

501 TGTGACAAAG CCTGTCGCCC GCCAGTTGCC CCTGGAGGGA AGTACTAAGT

551 AAAACTCAAT CCTGTCTTAA AGTGTGGCTG CAGGGGCCAG
AGGAGAGCCA

601 GCACGCACCA TGGCACCGGC CGGACACATC CTCACCTTGC
TGCTCTGGGG

651 TCACCTGCTG GAACTCTGGA CCCCAGGTCA CTCCGCGAAC
CCCTCCTACC

701 CCAGGCTACG CCTGTCACAT AAAGAACTTT TGGAACTGAA
TAGGACTTCA

751 ATATTTCAAA GCCCCCTTGG ATTTCTTGAT CTCCATACAA
TGCTGCTGGA

801 TGAGTATCAA GAACGGCTCT TTGTGGGAGG CAGAGACCTT
GTCTATTCCC

851 TGAACTTGGA ACGAGTCAGT GACGGCTACA GAGAGATATA
CTGGCCGAGC

901 ACAGCAGTAA AGGTAGAAGA ATGCATAATG AAAGGAAAAG
ACGCAAATGA

951 GTGTGCCAAT TATATCCGGG TTTTGCATCA CTACAACAGG
ACACACCTTC

1001 TGACCTGTGC TACTGGAGCT TTTGATCCAC ACTGTGCCTT
CATCAGAGTC

1051 GGGCACCATT CAGAGGAACC CCTGTTTCAC CTGGAGTCAC
ACAGATCTGA

1101 GAGAGGAAGG GGCAGATGTC CTTTTGACCC CAACTCCTCC
TTTGTGTCCA

1151 CGCTAGTTGG GAATGAGCTG TTTGCTGGAC TCTACAGTGA CTATTGGGGC

1201 AGAGACTCGG CGATCTTCCG CAGCATGGGG AAGTTAGGCC ATATTCGCAC

1251 TGAGCATGAC GATGAGCGGC TCCTGAAAGA ACCAAAATTT GTAGGTTCAT

1301 ATATGATTCC TGATAACGAA GACCGAGATG ACAACAAAAT GTACTTTTTC

1351 TTTACTGAGA AGGCGCTGGA GGCGGAGAAC AACGCCCACA CGATCTACAC

1401 CCGAGTGGGG CGGCTGTGCG TGAATGACAT GGGAGGACAG AGAATCCTGG

1451 TGAACAAGTG GAGCACTTTC CTTAAAGCGC GGCTGGTTTG CTCAGTGCCG

1501 GGAATGAATG GAATCGACAC ATACTTTGAC GAACTAGAGG ATGTGTTTTT

1551 ACTGCCGACC AGAGATCCTA AGAATCCAGT GATATTTGGA CTGTTTAATA

1601 CTACCAGCAA TATATTTAGA GGCCATGCTG TATGTGTGTA TCACATGTCA

1651 AGTATCCGGG AAGCCTTTAA TGGCCCATAT GCTCATAAAG AAGGCCCTGA

1701 ATACCACTGG TCACTATATG AAGGAAAAGT CCCCTACCCA AGGCCTGGTT

1751 CCTGTGCCAG CAAAGTAAAC GGAGGCAAGT ATGGAACCAC
CAAAGATTAC

1801 CCCGATGACG CCATCCGGTT CGCAAGGATC GATCCTCTAA
TGTATCAGCC

1851 CATAAAACCT GTTCATAAAA AACCAATACT GGTAAAAACA
GATGGAAAAT

1901 ACAACCTGAG GCAACTTGCC GTGGATCGGG TGGAAGCGGA
GGATGGCCAG

1951 TATGACGTCT TATTTATTGG GACAGACACA GGAATTGTGC
TGAAAGTAAT

2001 CACAATTTAC AACCAAGAAA CAGAGTGGAT GGAGGAAGTC
ATTCTAGAGG

2051 AACTTCAAAT ATTCAAGGAT CCAGCCCCTA TCATTTCTAT
GGAAATTTCT

2101 TCAAAGAGAC AACAGCTTTA CATTGGATCA GCCTCTGCTG
TGGCACAAGT

2151 CAGATTCCAT CACTGCGACA TGTATGGCAG TGCTTGTGCT
GACTGCTGCC

2201 TGGCTCGAGA CCCGTACTGT GCCTGGGATG GCATATCCTG
CTCCAGGTAC

2251 TACCCAACAG GTGCACACGC AAAGAGGAGG TTCCGCAGGC
AGGACGTTCG

2301 GCATGGCAAC GCCGCCCAAC AGTGCTTTGG ACAGCAATTT
GTTGGAGACG

2351 CGTTGGACAG GACTGAAGAG AGGCTGGCTT ATGGCATAGA
GAGCAACAGT

2401 ACTCTGTTGG AATGCACCCC GCGATCACTA CAAGCAAAAG
TCATCTGGTT

2451 TGTACAGAAG GGACGCGACG TAAGAAAAGA AGAGGTGAAG
ACGGATGACA

2501 GAGTTGTCAA GATGGACTTG GGCTTGCTCT TCCTCAGAGT
ACGCAAGTCA

2551 GATGCAGGGA CCTATTTTTG CCAGACAGTA GAACACAATT
TTGTCCATAC

2601 TGTGCGTAAA ATCACCTTGG AGGTGGTCGA AGAGCATAAA
GTGGAGGGCA

2651 TGTTTCATAA GGACCATGAA GAGGAAAGAC ATCACAAGAT
GCCCTGCCCT

2701 CCCTTAAGCG GTATGTCTCA GGGGACAAAA CCGTGGTACA
AGGAATTCTT

2751 GCAGCTGATT GGCTACAGCA ACTTCCAGAG AGTGGAAGAA
TACTGCGAAA

2801 AGGTGTGGTG TACAGATAAG AAGAGGAAAA AGCTTAAAAT
GTCTCCCTCC

2851 AAGTGGAAGT ATGCCAACCC CCAGGAAAAG AGGCTTCGCT
CTAAAGCTGA

2901 GCACTTCCGC CTGCCCAGGC ACACGCTGCT CTCCTGAGGG
CGCCCTCTGC

2951 CGGCGGCTGA GGAACCTAGG ATGGAAACAT TTTTTTAAAG
GGGGGGAAAA

3001 AAAAAAGACT GAAAGCATGC AGTCTCTTTC ATTACTTCAA
AGAACTTTCT

3051 GTAGTACTGA GAGGCTGGGA AGGTGTTTTA AGTTATTCTG
CATATCATCT

3101 GACTGTACAA ACATCTCTCT CACAGTACTA GTACTTAAAC
GAGTGTATTG

3151 CTTTAAAAGG CTTCACCTAT TTATCAACAT ATGATCGGGC
TGTAATTTTG

3201 AGCTTGCTTT TCAGGGAAAT TATTTTGCCA TTGGGATATG
TATTATAATT

3251 ATAATATTAA AGAAAAAGTT ATAGGTTTTC ATAATCAAGA
AAAATGTAAA

3301 ACTCTTTCTG GAGTTTTTTC TTTTTTCTTT TCCTCTTTGC
AGCCATAAGG

3351 TAAAATATAC AGAATATCAG ATATTGTAGA ATTATCGTGT
GAGCTTTCTG

3401 GTCAGTTGTT GATTTTTTTT TTTAAGCCAT AAGTTGTATG
TAGCTGTGTT

3451 GTGGTTGCTA TAAACTGTAC AAGAAATCTG CAACCCATCC
ATTCTGAGGC

3501 TTGTTCTCTT CATGAAACAC TCACTAAACT ATATTAAGAA
TATTTCTGTT

3551 TTATAAACAA GACATATCTT TTAAATTTCA CTTCAGTGTG
GAAGGATGGA

3601 AGATAGCCTC TCAGAAAGAT ATTTTTAAAT AGTTTCAAAC
CATGAAAGTA

3651 GAATTCTGTT AAAAACGGTC TGTGCTTCCC TCTTGTGGTA
CAAGAAGGCA

3701 TTCGATCTCT CAATGCTGTT TCCATAATTT TCCTCGAATG
ACTTGAAAAC

3751 AAAATGTAGT AGATACATCT AGTTCACTAT TCAGAAGACT
CAATTAAAAA

3801 ATAATTTTAT ATAATAATTT AATATGGAAA TAGCATGCTT
CTGGTTAATT

3851 CATTACTTCA CAACAAACCG AAGATGTTAA CTGCATTTGA
GCTTTGAAAT

3901 AGAAGAAAGT ACCTTTCCAG TCCTACCCTG AGACAAGACA
CTCTGAAAGT

3951 GTGCTTTACC CGCAAAGAAA TTATCAACGA TTCTTTTCTC
AGAAATGAAA

4001 TCTTCCTATT TTATATTTAC TTCTTGGATT CGTTTTACTA
CTGACATCTT

4051 TAGAGAATCA TTATCTTTGT CTTATGAATA TAATCAAAAT
TTATATTTTT

4101 CAGAATGTTT ATTTCTTGAG ATTTTTCTTC TCTGTATTTT
AAAGAGCCAG

4151 CAGCTTCCGT TCTAAAATAT GATTTTTATG AGGCTTTTCA
GTTCGGTAAG

4201 AGTAACGCAA GTGGATTGTA ATAAGAACAA TTATGAATAC
GTATTCAGAC

4251 TTAAATTTAA CTATTATGTC AGCTATCTCA GTATGCTATA
TAATTAAGAA

4301 TATAAAGTGT TGAGAACTTC TAATGGATTT CTTTTTCTGT
CTAAAATTAG

4351 GTACTACATA AATATTCCAC AATTTTCAAA TTTTAAAATT
TTGGAAATTT

4401 TAAAGTCCAG AATTTCTAAA ACATATAACT AGTATGAATA
AAGAGAATGT

4451 TGACATCCTC

SEQ ID NO: 2 SemaHv DNA sequence

1 AATTCGGCAC GAGTTCCTAG AAACGCGCGC GGGCTCAACC CTGCCTGAAC

51 TTTCCTGTAA ACAGAGCGCT GACAGGCGGC ATCCCCGCTG GGTGGATCCC

101 GCGCCCTGGC GCAAGTGGCA CTTCTTGCTT CTAATTATCG AGAGGAGAGG

151 CGAATACGAA CTAGCTGCTC GGCAAGTCAG TGTCAGGAGG CTGACTTCTG

201 GGAGGCTGGC GGGGAGGCTG GGGGAAGAGC TGGGGGAGGC TGCTGCTCTG

251 CTTCGACTGT TTTCTCAATG AATAGCTGGC GGGGAGACTG AAGCTAGCCA

301 CAGCCTCCTC CTTCACTCCG CGTCTGGGCT GACGGCGACA GCACCAGCCC

351 GGACCTGGCT CTCAAGACGC GCTCCTTGGA CGGTCTCTTG CTCCGCGCTT

401 CTAACCACCG GGCCCAAAGA CAGAAAGGCT TAGCGGATCC AAATATTGCC

451 CGGCAAATGG CACTTGGGAA TGGTATTTTC TGATGACAAC CCCTTCTGTT

501 TGTGACAAAG CCTGTCGCCC GCCAGTTGCC CCTGGAGGGA AGTACTAAGT

551 AAAACTCAAT CCTGTCTTAA AGTGTGGCTG CAGGGGCCAG AGGAGAGCCA

601 GCACGCACCA TGGCACCGGC CGGACACATC CTCACGTTGC TGCTCTGGGG

651 TCACCTGCTG GAACTCTGGA CCCCAGGTCA CTCGGCGAAC CCCTCCTACG

701 CCAGGCTACC GCTGTCACAT AAAGAACTTT TCGAACTGAA TGGACTTCAA

751 TATTTCAAAG CCCCCCTTGG ATTTCTTGAT CTCCATACAA TGCTGCTGGA

801 TGAGTATCAA GAACGGCTCT TTGTGGGAGG CAGAGACCTT GTCTATTCCC

851 TGAACTTGGA ACGAGTGAGT GACGGCTACA GAGAGATATA CTGGCCGAGC

901 ACAGCAGTAA AGGTAGAAGA ATGCATAATG AAAGGAAAAG ACGCAAATGA

951 GTGTGCCAAT TATATCCGGG TTTTGCATCA CTACAACAGG ACACACCTTC

1001 TGACCTGTGC TACTGGTGCT TTTGATCCAC ACTGTGCCTT
CATCAGAGTC

1051 GGGCACCATT CAGAGGAACC CCTGTTTCAC CTGGAGTCAC
ACAGATCTGA

1101 GAGAGGAAGG GGCAGATGTC CTTTTGACCC CAACTCCTCC
TTTGTGTCCA

1151 CGCTAGTTGG GAATGAGCTG TTTGCTGGAC TCTACAGTGA
CTATTGGGGC

1201 AGAGACTCGG CGATCTTCCG CAGCATGGGG AAGTTAGGCC
ATATTCGCAC

1251 TGAGCATGAC GATGAGCGGC TCCTGAAAGA ACCAAAATTT
GTAGGTTCAT

1301 ATATGATTCC TGATAACGAA GACCGAGATG ACAACAAAAT
GTACTTTTTC

1351 TTTACTGAGA AGGCGCTGGA GGCGGAGAAC AACGCCCACA
CGATCTTACA

1401 CCCGAGTGGG CGGCTGTGCG TGAATGACAT GGGAGGACAG
AGAATCCTGG

1451 TGAACAAGTG GAGCACTTTC CTTAAAGCGC GGCTGGTTTG
CTCAGTGCCG

1501 GGAATGAATG GAATCGACAC ATACTTTGAC GAACTAGAGG
ATGTGTTTTT

1551 ACTGCCGACC AGAGATCCTA AGAATCCAGT GATATTTGGA
CTGTTTAATA

1601 CTACCAGCAA TATATTTAGA GGCCATGCTG TATGTGTGTA
TCACATGTCA

1651 AGTATCCGGG AAGCCTTTAA TGGCCCATAT GCTCATAAAG
AAGGCCCTGA

1701 ATACCACTGG TCACTATATG AAGGAAAAGT CCCCTACCCA
AGGCCTCGTT

1751 CCTGTGCCAG CAAAGTAAAC GGAGGCAAGT ATGGAACCAA
CCAAAGATTA

1801 CCCGATGACG CCATCCGGTT CGCAAGGATG CATCCTCTAA
TGTATCAGCC

1851 CATAAAACCT GTTCATAAAA AACCAATACT GGTAAAAACA
GATGGAAAAT

1901 ACAACCTGAG GCAACTTGCC GTGGATCGGG TGGAAGCGGA
GGATGGCCAG

1951 TATGACGTCT TATTTATTGG CACAGACACA GGAATTGTGC
TGCTGAAAGT

2001 AATCACAATT TACAACCAAG AAACAGAGTG GATGGAGGAA
GTCATTCTAG

2051 AGGAACTTCA AATATTCAAG GATCCAGCCC CTATCATTTC
TATGGAAATT

2101 TCTTCAAAGA GACAACAGCT TTACATTGGA TCAGCCTCTG
CTGTGGCACA

2151 AGTCAGATTC CATCACTGCG ACATGTATGG CAGTGCTTGT
GCTGACTGCT

2201 GCCTGGCTCG AGACCCGTAC TGTGCCTGGG ATGGCATATC
CTGCTCCAGG

2251 TACTACCCAA CAGGTGCACA CGAAAAGAGG AGGTTCCGCA
GGCAGGACGT

2301 TCGGCATGGC AACGCCGCCC AACAGTGCTT TGGACAGCAA
TTTGTTGGAG

2351 ACGCGTTGGA CAGGACTGAA GAGAGGCTGG CTTATGGCAT
AGAGAGCAAC

2401 AGTACTCTGT TGGAATGCAC CCCGCTATCA CTACAAGCAA
AAGTCATCTG

2451 GTTTCTACAG AAGGGACGCG ACGTAAGAAA AGAAGAGGTG
AAGACGGATG

2501 ACAGAGTTGT CAAGATGGAC TTGGGCTTGC TCTTCCTCAG
AGTACGCAAG

2551 TCAGATGCAG GGACCTATTT TTGCCAGACA GTAGAACACA
ATTTTGTCCA

2601 TACTGTGCGT AAAATCACCT TGGAGGTGGT CGAAGAGCAT
AAAGTGGAGG

2651 GCATGTTTCA TAAGGACCAT GAAGAGGAAA GACATCACAA
GATGCCCTGC

2701 CCTCCCTTAA GCGGTATGTC TCAGGGGACA AAACCGTGGT
ACAAGGAATT

2751 CTTGCAGCTG ATTGGCTACA GCAGCAAGTT CCAGAGAGTG
GAAGAATACT

2801 GCGAAAAGGT GTGGTGTACA GATAAGAAGA GGAAAAAGCT
TAAAATGTCT

2851 CCCTCCAAGT GGAAGTATGC CAACCCCCAG GAAAAGAGGC
TTCGCTCTAA

2901 AGCTGAGCAC TTCCGCCTGC CCAGGCACAC GCTGCTCTCC
TGAGGGCGCC

2951 CTCTGCCGGC GGCTGAGGAA CCTAGGATGG AAACATTTTT
TTAAAGGGGG

3001 GGAAAAAAAA AAGACTGAAA GCATGCAGTC TCTTTCATTA
CTTCAAAGAA

3051 CTTTCTGTAG TACTGAGAGG CTGGGAAGGT GTTTTAAGTT
ATTCTGCATA

3101 TCATCTGACT GTACAAACAT CTCTCTCACA GTACTAGTAC
TTAAACGAGT

3151 GTATTGCTTT AAAAGGCTTC ACCTATTTAT CAACATATGA
TCGGGCTGTA

3201 ATTTTGAGCT TGCTTTTCAG GGAAATTATT TTGCCATTGG
GATATGTATT

3251 ATAATTATAA TATTAAAGAA AAAGTTATAG GTTTTCATAA
TCAAGAAAAA

3301 TGTAAAACTC TTTCTGGAGT TTTTTCTTTT TTCTTTTCCT
CTTTGCAGCC

3351 ATAAGGTAAA ATATACAGAA TATCAGATAT TGTAGAATTA
TCGTGTGAGC

3401 TTTCTGGTCA GTTGTTGATT TTTTTTTTTA AGCCATAAGT
TGTATGTAGC

3451 TGTGTTGTGG TTGCTATAAA CTGTACAAGA AATCTGCAAC
CCATCCATTC

3501 TGAGGCTTGT TCTCTTCATG AAACACTCAC TAAACTATAT
TAAGAATATT

3551 TCTGTTTTAT AAACAAGACA TATCTTTTAA ATTTCACTTC
AGTGTGGAAG

3601 GATGGAAGAT AGCCTCTCAG AAAGATATTT TTAAATAGTT
TCAAACCATG

3651 AAAGTAGAAT TCTGTTAAAA ACGGTCTGTG CTTCCCTCTT
GTGGTACAAG

3701 AAGGCATTCG ATCTCTCAAT GCTGTTTCCA TAATTTTCCT
CGAATGACTT

3751 GAAAACAAAA TGTAGTAGAT ACATCTAGTT CACTATTCAG
AAGACTCAAT

3801 TAAAAAATAA TTTTTATATA ATTAAGAATA TAAAGTGTTG
AGAACTTCTA

3851 ATGGATTTCT TTTTCTGTCT AAAATTAGGT ACTACATAAA
TATTCCACAA

3901 TTTTCAAATT TTAAAATTTT GGAAATTTTA AAGTCCAGAA
TTTCTAAAAC

3951 ATATAACTAG TATGAATAAA GAGAATGTTG ACATCCTC

SEQ ID NO: 3 Sema-H amino acid sequence

MAPAGHILTLLLWGHLLELWTPGHSANPSYPRLRLSHKELLELN

RTSIFQSPLGFLDLHTMLLDEYQERLFVGGRDLVYSLNLERVSDGYREIYWPSTAVKV

EECIMKGKDANECANYIRVLHHYNRTHLLTCATGAFDPHCAFIRVGHHSEEPLFHLES

HRSERGRGRCPFDPNSSFVSTLVGNELFAGLYSDYWGRDSAIFRSMGKLGHIRTEHDD

ERLLKEPKFVGSYMIPDNEDRDDNKMYFFFTEKALEAENNAHTIYTRVGRLCVNDMGG

QRILVNKWSTFLKARLVCSVPGMNGIDTYFDELEDVFLLPTRDPKNPVIFGLFNTTSN

IFRGHAVCVYHMSSIREAFNGPYAHKEGPEYHWSLYEGKVPYPRPGSCASKVNGGKYG

TTKDYPDDAIRFARIDPLMYQPIKPVHKKPILVKTDGKYNLRQLAVDRVEAEDGQYDV

LFIGTDTGIVLKVITIYNQETEWMEEVILEELQIFKDPAPIISMEISSKRQQLYIGS

ASAVAQVRFHHCDMYGSACADCCLARDPYCAWDGISCSRYYPTGAHAKRRFRRQDVRH

GNAAQQCFGQQFVGDALDRTEERLAYGIESNSTLLECTPRSLQAKVIWFYQKGRDVRK

EEVKTDDRVVKMDLGLLFLRVRKSDAGTYFCQTVEHNFVHTVRKITLEVVEEHKVEGM

FHKDHEEERHHKMPCPPLSGMSQGTKPWYKEFLQLIGYSNFQRVEEYCEKVWCTDKK

RKKLKMSPSKWKYANPQEKRLRSKAEHFRLPRHTLLS

SEQ ID NO: 4 SemaHv amino acid sequence

MAPAGHILTLLLWGHLLELWTPGHSANPSYARLPLSHKELFELN

GLQYFKAPLGFLDLHTMLLDEYQERLFVGGRDLVYSLNLERVSDGYREIYWPSTAVKV

EECIMKGKDANECANYIRVLHHYNRTHLLTCATGAFDPHCAFIRVGHHSEEPLFHLES

HRSERGRGRCPFDPNSSFVSTLVGNELFAGLYSDYWGRDSAIFRSMGKLGHIRTEHDD

ERLLKEPKFVGSYMIPDNEDRDDNKMYFFFTEKALEAENNAHTILHPSGRLCVNDMGG

QRILVNKWSTFLKARLVCSVPGMNGIDTYFDELEDVFLLPTRDPKNPVIFGLFNTTSN

IFRGHAVCVYHMSSIREAFNGPYAHKEGPEYHWSLYEGKVPYPRPRSCASKVNGGKYG

TNQRLPDDAIRFARMHPLMYQPIKPVHKKPILVKTDGKYNLRQLAVDRVEAEDGQYDV

LFIGTDTGIVLLKVITIYNQETEWMEEVILEELQIFKDPAPIISMEISSKRQQLYIGS

ASAVAQVRFHHCDMYGSACADCCLARDPYCAWDGISCSRYYPTGAHEKRRFRRQDVRH

GNAAQQCFGQQFVGDALDRTEERLAYGIESNSTLLECTPLSLQAKVIWFLQKGRDVRK

EEVKTDDRVVKMDLGLLFLRVRKSDAGTYFCQTVEHNFVHTVRKITLEVVEEHKVEGM

FHKDHEEERHHKMPCPPLSGMSQGTKPWYKEFLQLIGYSSKFQRVEEYCEKVWCTDKK

RKKLKMSPSKWKYANPQEKRLRSKAEHFRLPRHTLLS